

Finding Weighted Graphs by Combinatorial Search

Jeong Han Kim*

Abstract

We consider the problem of finding edges of a hidden weighted graph using a certain type of queries. Let G be a weighted graph with n vertices. In the most general setting, the n vertices are known and no other information about G is given. The problem is finding all edges of G and their weights using additive queries, where, for an additive query, one chooses a set of vertices and asks the sum of the weights of edges with both ends in the set. This model has been extensively used in bioinformatics including genom sequencing. Extending recent results of Bshouty and Mazzawi [11], and Choi and Kim [17], we present a polynomial time randomized algorithm to find the hidden weighted graph G when the number of edges in G is known to be at most $m \geq 2$ and the weight $w(e)$ of each edge e satisfies $\alpha \leq |w(e)| \leq \beta$ for fixed constants $\alpha, \beta > 0$. The query complexity of the algorithm is $O(\frac{m \log n}{\log m})$, which is optimal up to a constant factor.

The algorithm heavily relies on a well-known combinatorial search problem, which may be of independent interest. Suppose that there are n identical looking coins and some of them are counterfeit. The weights of all authentic coins are the same and known a priori. The weights of counterfeit coins vary but different from the weight of an authentic coin. Without loss of generality, we may assume the weight of authentic coins is 0. The problem is to find all counterfeit coins by weighing sets of coins on a spring scale. We introduce a polynomial time randomized algorithm to find all counterfeit coins when the number of them is known to be at most $m \geq 2$ and the weight $w(c)$ of each counterfeit coin c satisfies $\alpha \leq |w(c)| \leq \beta$ for fixed constants α, β . The query complexity of the algorithm is $O(\frac{m \log n}{\log m})$, which is optimal up to a constant factor. The algorithm uses, in part, random walks.

Keywords – graph finding, combinatorial search, coin weighing, additive query, random walk

1 Introduction

1.1 Graph Finding Problem

We consider the problem of finding edges of a hidden weighted graph using a certain type of queries. Let G be a weighted graph with n vertices. In the most general setting, the n vertices are known and no other information about G is given. The problem is finding all edges of G and their weights using queries. Three types of queries have been extensively studied:

Detection query: One chooses a set of vertices and asks if there is an edge with both ends in the set. This type of queries has applications to genom sequencing and has been studied in [1, 2, 3, 4, 24, 25].

Additive query: One chooses a set of vertices and asks the sum of weights of edges with both ends in the set. This model has been extensively used in bioinformatics including genom sequencing, and studied in [3, 6, 8, 9, 10, 11, 18, 23, 25, 26, 36, 39].

Shortest path query: One choose a pair of vertices and asks the length of the shortest path between the two vertices. This query arises in the canonical model of the evolutionary tree literature [27, 29, 40].

(Our lists of references are far from being exhaustive.)

In this paper, we focus on the additive queries. The graph finding problem with additive queries is partly motivated by the shotgun sequencing [5, 25], one of the most popular methods for DNA sequencing.

*Department of Mathematics, Yonsei University, Seoul, 120-749 Korea (e-mail: jehkim@yonsei.ac.kr).

In the shotgun sequencing, one needs to put back separately decoded short fragments of a given genome sequence into the same order as in the original sequence. Combined with a biotech method called the multiplex PCR [43], the process is reduced to the problem of finding a hidden graph using additive queries. The additive queries are also used in the problem of finding the Fourier coefficients of pseudo-Boolean functions, which play crucial roles in evolutionary computation, artificial intelligence, and population genetics [18, 16, 19].

In the rest of this paper, we say queries for additive queries and all logarithms are in base 2, unless otherwise specified. For unweighted graphs, Grebinski and Kucherov presented a few results. For arbitrary graphs on n vertices, they have shown that $O(\frac{n^2}{\log n})$ queries are enough [26]. If the hidden graph is known to be a Hamiltonian path or cycle, then $O(n)$ queries are suffice [25]. More generally, if the maximum degree of the hidden graph is known to be at most d , then the graph may be found using $O(dn)$ queries [26]. Grebinski [23] has shown that the same bound $O(dn)$ holds for d -degenerate graphs.

When the hidden graph has at most $m \geq 2$ edges and m is known, some bounds close to the optimal bound were shown [3, 39] and Choi and Kim [18] proved a $O(\frac{m \log(n^2/m)}{\log m})$ bound that is optimal (up to a constant factor). The randomized algorithm presented there uses non-adaptive queries but it is not a polynomial time algorithm, where queries are non-adaptive if each query is independent of answers to the previous queries. Recently, Mazzawi [36] constructed a polynomial time algorithm with optimal query complexity. The algorithm is deterministic and uses adaptive queries. She also extended the algorithm to find weighted graphs with positive integer weights.

For weighted graphs, Choi and Kim [18] proved a non-adaptive $O(\frac{m \log n}{\log m})$ query bound, provided that m is at least a polylog of n and the absolute values of all weights are between n^{-a} and n^b for constants $a, b > 0$. Bshouty and Mazzawi [11] showed the same bound without the extra conditions. However, it is unlikely that one may able to develop a polynomial time algorithm from those results. In other words, substantially new ideas seem to be needed to design an algorithm that is useful in practical sense. A significant result toward this direction has been shown by Bshouty and Mazzawi [9]: For weighted graphs with positive real weights, they presented a deterministic polynomial time algorithm that uses an almost optimal number of (adaptive) queries, $O(\frac{m \log n}{\log m} + m \log \log m)$. Note that the extra $m \log \log m$ term is larger than the optimal query bound by a $\log \log n$ factor when $\log m = \Omega(\log n)$.

To obtain the optimal query complexity $O(\frac{m \log n}{\log m})$, Choi and Kim [17] have recently introduced a randomized polynomial time algorithm that finds the hidden weighted graph with positive real weights. Another randomized polynomial time algorithm they introduced uses $O(\frac{m \log n}{\log m})$ queries to find the hidden weighted graph with bounded integer weights.

In this paper, we present a randomized polynomial time algorithm that works for a quite general class of weighted graphs. Using the optimal number of queries up to constant factor, the algorithm finds the hidden weighted graph provided that the weight $w(e)$ of each edge e in the graph satisfies $\alpha \leq |w(e)| \leq \beta$ for positive constants α and β . The theorem we will prove is slightly more general in the sense that α, β are not necessarily constants.

Theorem 1.1. *Let n, m be positive integers with $n^2 \geq m \geq 2$ and let $\alpha, \beta > 0$ be positive real numbers (not necessarily constants) with $2\alpha < \beta$. Suppose a weighted graph G with n vertices and at most m edges is given. If the weight $w(e)$ of each edge in G satisfies $\alpha \leq |w(e)| \leq \beta$, then there is a randomized polynomial time algorithm that asks $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries to find all edges with probability $1 - O(1/m^{0.02})$.*

Our proof of the theorem heavily relies on a well-known combinatorial search problem. Suppose there are n identical looking coins and some of them are counterfeit. The weights of all authentic coins are the same and known a priori. The weights of counterfeit coins vary but different from the weight of an authentic coin. The problem is to find all counterfeit coins by weighing sets of coins on a spring scale. Note that weighing sets of coins on a spring scale may be regarded as additive queries. This problem is also equivalent to the graph finding problem when the graphs are restricted to stars $K_{1,m}$ with known center. The coin weighing problem has been extensively studied. We survey its colorful history and add one more algorithm finding all counterfeit coins when the weights of each counterfeit coin satisfies properties similar to those described in the above theorem.

1.2 Coin Weighing Problem

Suppose there are n identically looking coins, some of them are counterfeit. The weights of all authentic coins are the same and known a priori, while the weights of counterfeit coins are unknown but different from the weight of an authentic coin. Without loss of generality, it may be assumed that the weights of authentic coins are 0 and the weights of counterfeit coins belong to a set of non-zero real numbers. We want to find all counterfeit coins by weighing sets of coins on a spring scale, which we call additive queries or simply queries.

After the coin weighing problem was introduced by Fine [22] and Shapiro [41], a number of results have been published, mainly focusing on the case that the weights of counterfeit coins are the same [12, 13, 21, 31, 32, 33, 38, 42]: Summarizing some of them briefly, Erdős and Rényi [21], in 1963, proved that $\frac{(\log 9 + o(1))n}{\log n}$ queries are enough and $\frac{(2+o(1))n}{\log n}$ queries are required. (See [30] for another proof of the lower bound.) The upper bound was improved to match the lower bound by Cantor and Mills [13], and Lindström [32]. Using the Möbius function, Lindström [33, 34] explicitly constructed a query matrix that asks $\frac{(2+o(1))n}{\log n}$ queries. The case that the number m of counterfeit coins is also known has been extensively studied too [14, 15, 20, 26, 34, 35, 44, 45]. Recently, Bshouty [7] proposed the first polynomial time algorithm that uses $\frac{(1+o(1))2m \log \frac{n}{m}}{\log m}$ adaptive queries. The query complexity is optimal up to $o(1)$ term.

Results for the general case, in which the weights of counterfeit coins are not the same, have been obtained only recently. As the results were applied to the (weighted) graph finding problem, our summary is almost the same as in the previous subsection. When the weights of the counterfeit coins can be any (not necessarily positive) real numbers, Choi and Kim [18] proposed an algorithm with a non-adaptive $O(\frac{m \log n}{\log m})$ query bound, under the mild conditions on m and the weights, i.e., $m = \Omega(\text{polylog } n)$ and the absolute values of all weights are between n^{-a} and n^b for constants $a, b > 0$. Bshouty and Mazzawi [11] showed the same bound without the extra conditions. Though the query complexities of both algorithms are optimal, the time complexities of them are far from being polynomial. Concerning polynomial time algorithms, Bshouty and Mazzawi [9] presented a deterministic polynomial time algorithm that uses a near optimal number of (adaptive) queries, $O(\frac{m \log n}{\log m} + m \log \log m)$, assuming the weights of all counterfeit coins are positive real numbers. They first constructed a search matrix using Fourier representations, and took the divide and conquer approach to guess the sums of the weights of coins. The search matrix played key roles when the sums of the weights were guessed. The processes for checking and correction follow after guessing.

As mentioned before, the extra $m \log \log m$ term is larger than the optimal bound by a $\log \log n$ factor when $\log m = \Omega(\log n)$. Choi and Kim [17] presented a polynomial time randomized algorithm to remove the $m \log \log m$ term in the query complexity. Another polynomial time randomized algorithm may be applied to achieve the optimal query complexity, when the weights of counterfeit coins are bounded integers in absolute values. The key idea is constructing random sets of coins that are useful to control the number of checking and correction processes used by Bshouty and Mazzawi [9]. Once the number of checking and correction processes is substantially reduced, less queries are needed.

A randomized algorithm is presented in this paper to achieve the optimal query complexity when the weights of counterfeit coins are any real numbers bounded from below and from above in absolute values. The theorem we will prove is slightly more general in the sense that some exceptions for the weight condition are allowed.

Theorem 1.2. *Let n, m be positive integers with $n \geq m \geq 2$ and let $\alpha, \beta, \varepsilon > 0$ be positive real numbers (not necessarily constants) with $2\alpha < \beta, \varepsilon < 1/2$. Suppose n coins are given and there are at most m counterfeit coins among them. The weights of authentic coins are 0 while the weights of counterfeit coins vary but they are non-zero. If the weights $w(c)$ of all but εm counterfeit coins c satisfy $\alpha \leq |w(c)| \leq \beta$ and the weights $w(c)$ of the εm counterfeit coins c satisfies just $|w(c)| \leq \beta$, then there is a randomized polynomial time algorithm that asks $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries and finds all but $m^{0.8} + 2\varepsilon m$ counterfeit coins, with probability $1 - O(1/m^{0.8})$. All the remaining counterfeit coins can be found using $O((m^{0.8} + 2\varepsilon m) \log n)$ additional queries, with probability $1 - e^{-\Omega(m^{0.8})}$.*

In the proof of Theorem 1.2, we use the search matrix Bshouty and Mazzawi [9] developed after constructing random sets of coins as in Choi and Kim [17]. Though the guessing processes are the same as in [17], the processes for checking and correction are newly developed using biased random walks.

One may easily verify if the coins declared to be counterfeit by the algorithm in Theorem 1.2 are actually counterfeit by directly weighing, using m additional queries. Running the algorithm $O(\mu)$ times with the verification at each time, the error probability may be made arbitrarily small.

Corollary 1.3. *Under the same hypotheses of Theorem 1.2 and any integer $\mu \geq 1$, there is a randomized polynomial algorithm that uses $O(\frac{\mu m \log(\beta/\alpha) \log n}{\log m})$ queries and finds all but $m^{0.8} + 2\epsilon m$ counterfeit coins with probability $1 - O(1/m^\mu)$. All the remaining counterfeit coins can be found using $O((m^{0.8} + 2\epsilon m) \log n)$ additional queries, with probability $1 - e^{-\Omega(m^{0.8})}$.*

After presenting the search matrix and two martingale inequalities in Section 2, we prove Theorem 1.2 in Section 3. Section 4 is for the proof of Theorem 1.1. The concluding remark will follow.

2 Preliminaries

As mentioned in the previous section, Bshouty and Mazzawi [9] used Fourier representation of certain functions to find a search matrix, i.e., a 0,1 matrix that is useful for coin weighing problems. We present properties of the matrix in a slightly generalized form.

Lemma 2.1. *Let γ, m be positive integers. Then, for the smallest integer t satisfying $t2^{t-1} \geq \gamma m$, one can construct, in polynomial time, $2^t \times m$ 0,1 matrix S and $2^t \times 2^t$ matrix T with the following property: For each $j = 1, \dots, m$, one may find, in polynomial time, a unique positive integer $i_j \leq 2^t$ and a non-negative integer $k_j \leq \lceil t/\gamma \rceil - 1$ satisfying*

$$(TS)_{ijk} = 2^{-(k-j)\alpha} (TS)_{ijj} \text{ for } j+1 \leq k \leq j+k_j, \text{ and } (TS)_{ijk} = 0 \text{ for } k \geq j+k_j+1,$$

where $(TS)_{ij}$ is the ij entry of TS .

Setting $a_{jk} = \frac{(TS)_{ijk}}{(TS)_{ijj}}$, we have the following corollary.

Corollary 2.2. *Let γ, m are positive integers and t be the smallest integer satisfying $t2^{t-1} \geq \gamma m$. Then one can find, in polynomial time, 2^t non-adaptive queries, real numbers a_{jk} , and a non-negative integer $k_j \leq \lceil t/\gamma \rceil - 1$, $j = 1, \dots, m$, $k = 1, \dots, j-1$, satisfying the following property: For disjoint sets A_1, \dots, A_m of coins, the 2^t queries yield values x_j , in polynomial time, satisfying*

$$w(A_j) = x_j - \sum_{k=1}^{j-1} a_{jk} w(A_k) - \sum_{k=1}^{k_j} \frac{w(A_{j+k})}{2^{k\alpha}},$$

$j = 1, \dots, m$, where $w(A)$ is the sum of weights of all coins in A . In particular, $\frac{(2+o(1))\gamma m}{\log(\gamma m)}$ queries are enough to find x_j 's.

We will need the Azuma-Hoeffding martingale inequality too. The following is from [37].

Lemma 2.3. *Let $Z = (Z_1, \dots, Z_t)$ be a family of independent random variables with Z_ℓ taking values in a finite set B_ℓ for each ℓ . Suppose that the real-valued function f defined on $\prod_\ell B_\ell$ satisfies*

$$|f(\mathbf{z}) - f(\mathbf{z}')| \leq c_\ell$$

whenever the vectors \mathbf{z} and \mathbf{z}' differ only in the ℓ^{th} coordinate. Then for any $\lambda \geq 0$,

$$\Pr [|f(Z) - \mathbb{E}[f(Z)]| \geq \lambda] \leq 2e^{-2\lambda^2 / \sum_\ell c_\ell^2}.$$

For our purpose, a more general martingale inequality is needed. The following version appeared in [28].

Lemma 2.4. *Let $X = (Z_1, \dots, Z_t)$ be independent identically distributed (i.i.d.) Bernoulli random variables with probability p (i.e., $\Pr[Z_i = 1] = p$ and $\Pr[Z_i = 0] = 1 - p$ for each i). Suppose that the real-valued function f defined on $\{0, 1\}^t$ satisfies*

$$|f(\mathbf{z}) - f(\mathbf{z}')| \leq c_i$$

whenever the vectors \mathbf{z} and \mathbf{z}' differ only in the i^{th} coordinate. Then for any $\lambda, \rho > 0$,

$$\Pr[|f(Z) - \mathbb{E}[f(Z)]| \geq \lambda] \leq 2 \exp\left(-\rho\lambda + (\rho^2/2)p(1-p) \sum_{i=1}^t c_i^2 \exp(\rho c_i)\right).$$

3 Coin Weighing Problem

Suppose n coins are given, some of which are counterfeit. The weights of all authentic coins are the same and known a priori, while the weights of counterfeit coins are unknown but different from the weight of an authentic coin. Without loss of generality, we may assume that the weights of authentic coins are 0 and the weights of counterfeit coins belong to a set of non-zero real numbers. We assume that the number of counterfeit coins is known to be at most m .

If $O(m \log n)$ queries are allowed to find counterfeit coins. One may use a randomized binary search:

Randomized Binary Search Suppose a set A of coins is given, and the number of coins is no more than n and there are at most $m \leq n$ counterfeit coins. Then select each coin with probability $1/2$, independently of all other coins. Then weigh the set A' of selected coins. If the weight is non-zero, then find a counterfeit coin among the selected coins, using the deterministic binary search.

The deterministic binary search is as follows. Divide A' into two parts A'_1, A'_2 with size difference at most 1. If $w(A'_1) \neq 0$, then select A'_1 . Otherwise, select A'_2 . Keep doing this for the selected set until a counterfeit coin is found.

Provided that there is a counterfeit coin, it is not hard to see that the probability of the weight of A' being non-zero is at least $1/2$ and the deterministic binary search requires no more than $\lceil \log n \rceil$ queries. The number of queries required to find one counterfeit coin is at most $2 + \lceil \log n \rceil$ in expectation. Thus, it is expected that $(\lceil \log n \rceil + 2 + o(1))m$ queries are enough to find all counterfeit coins, with high probability. Here, we show that $(\lceil \log n \rceil + 3)m$ queries are enough, with probability $1 - e^{-\Omega(m)}$.

Lemma 3.1. *With probability $1 - e^{-\Omega(m)}$, the randomized binary search finds all counterfeit coins using $(\lceil \log n \rceil + 3)m$ queries.*

The proof of the lemma is presented in Appendix.

We first construct random sets of coins and then present the algorithm, for which the time complexity is not optimized but it is clearly a polynomial time algorithm. Some explanation and analysis of the algorithm will follow after the algorithm is presented. The construction of random sets is the same as in Choi and Kim [17].

Constructing random sets of coins: Let A be a set of n or less coins. For an integer $q \geq 2$ and $\ell_q := \lceil \log q \rceil$, we construct random subsets $A_{i,j}$ of A , $i = 0, 1, \dots, \lceil 3 \log n \rceil$, $j = 1, \dots, 2^{\ell_q+i}$. For $i = 0$, we assign each coin in A a uniform random number among $1, \dots, 2^{\ell_q}$, independently of all other coins. The set $A_{0,j}$ consists of all coins with assigned number j . Generally, for $i = 1, \dots, \lceil 2 \log q \rceil - 1$, once all $A_{i-1,j}$, $j = 1, \dots, 2^{\ell_q+i-1}$, are constructed, we may randomly divide each set $A_{i-1,j}$ into two parts so that coins in $A_{i-1,j}$ are independently in the first part with probability $1/2$. The other coins in $A_{i-1,j}$ are to be in the second part. The set of all coins in the first and second parts are denoted by $A_{i,2j-1}$

and $A_{i,2j}$, respectively. Or equivalently, after assigning each coin mutually independent random numbers $r_0, r_1, \dots, r_{\lceil 2 \log q \rceil - 1}$, independently of all other coins, with

$$\Pr[r_0 = a] = 2^{-\ell_q}, \quad a = 1, \dots, 2^{\ell_q} \quad \text{and} \quad \Pr[r_i = a] = \frac{1}{2}, \quad a = 0, 1, \quad i = 1, \dots, \lceil 2 \log q \rceil - 1,$$

we define $A_{i,j}$ to be the set of all coins with assigned numbers $r_0, r_1, \dots, r_{\lceil 2 \log q \rceil - 1}$ satisfying $j = 1 + (r_0 - 1)2^i + r_1 2^{i-1} + \dots + r_i$.

For $i \geq \lceil 2 \log q \rceil$, $A_{i-1,j}$ may be deterministically divided into two parts so that the first part consists of $\lceil |A_{i-1,j}|/2 \rceil$ coins. As before, the first part is denoted by $A_{i,2j-1}$, and $A_{i,2j} = A_{i-1,j} \setminus A_{i,2j-1}$. This construction is expected to stop when all $A_{i,j}$, $j = 1, \dots, 2^{\ell_m+i}$, consist of one or no coin. As there are n coins, all $A_{i,j}$ consist of one or no coin within $\lceil \log n \rceil$ more rounds after the random construction ends. For the sake of safeness, we stop the construct when $i = \lceil 3 \log n \rceil \geq \lceil 2 \log q \rceil + \lceil \log n \rceil$.

The following lemma summarize properties of the random subsets $A_{i,j}$ that will be used for the analysis of the algorithm presented later. The proof is essentially in [17] and it is presented in Appendix for the sake of completeness.

Lemma 3.2. *Suppose a set A of n or less coins are given, and the number of counterfeit coins in A is at most $q \geq 2$. If the weights $w(c)$ of all but at most $q/2$ counterfeit coins c satisfy $|w(c)| \geq \alpha$. Then, with probability $1 - O(\frac{1}{q})$, we have the followings.*

- (a) *There are at most $\frac{5q}{6}$ counterfeit coins c that satisfy $|w(c)| < \alpha$ (not exclusive) or belong to a set $A_{0,j}$ containing more than one counterfeit coin, $j = 1, \dots, 2^{\ell_q}$.*
- (b) *For each $i = 1, \dots, \lceil 2 \log q \rceil - 1$, $A_{i,j}$ contains at most $\frac{i+2\log q}{i}$ counterfeit coins.*
- (c) *For each $i = 1, \dots, \lceil 2 \log q \rceil - 1$, there are at most $2^{-(i+1)}q + q^{3/4}$ sets $A_{i,j}$ that contain more than one counterfeit coin.*
- (d) *For $i \geq \lceil 2 \log q \rceil - 1$, each $A_{i,j}$ contains one or less counterfeit coin.*
- (e) *Each $A_{\lceil 3 \log n \rceil, j}$ contains at most one coin.*

Now we are ready to present the algorithm described in Theorem 1.2.

Algorithm (i) (Initially, $q = m$ and A is the set of all n coins.) Construct random subsets $A_{i,j}$ of A as above with parameter q . Then weigh $A_{0,j}$ for all $j = 1, \dots, 2^{\ell_q}$, and denote $w_{0,j} = w(A_{0,j})$ $j = 1, \dots, 2^{\ell_q}$ and J_0 to be the set of all j such that $|w_{0,j}| \geq \alpha$. Then go to (ii), where, in general, $w(B) = \sum_{c \in B} w(c)$ for a set B of coins.

(ii) (Initially $i = 1$ and $J = J_0$.) After relabeling, we may assume $J = \{1, \dots, |J|\}$. Apply Corollary 2.2 with $\gamma_i = \max\{\lceil \log(\frac{6\beta}{\alpha}) \rceil, \lceil \log(\frac{3\beta(i+2\log q)}{i\alpha}) \rceil\}$ to $A_{i,2}, \dots, A_{i,2|J|}$ and obtain x_r satisfying

$$w(A_{i,2r}) = x_r - \sum_{k=1}^{r-1} a_{rk} w(A_{i,2k}) - \sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}}. \quad (1)$$

Set, inductively in $r = 1, \dots, |J|$,

$$u_{2r} = \begin{cases} w_{i-1,r} & \text{if } |x_r - \sum_{k=1}^{r-1} a_{rk} u_{2k}| \geq \frac{\alpha}{2} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

and $u_{2r-1} = w_{i-1,r} - u_{2r}$, $r = 1, \dots, |J|$. Go to (iii) if $i < \lceil 2 \log q \rceil$. Otherwise, go to (iv).

(iii) (Initially, $s = -2$.) Randomly select each j satisfying $u_j = 0$ and $j \leq \min\{s, 2|J|\}$ with probability $1/2$, independently of all other j . Weigh $\cup\{A_{i,j} : \text{selected } j\}$. The weight is 0 if no j is selected. Do this random

weighing $\lceil \log(i^2 + 1) \rceil + 3$ times, independently of all other random weighings. This procedure is called a random test at s . If the test is passed, i.e., all weights are 0, then update s to be $s + 2i^2$. If it is failed and $s \leq 2|J|$, correct u_s by weighing $A_{i,s}$, that is, weigh $A_{i,s}$, and update u_s to be $w(A_{i,s})$ and u_{s-1} to be $w_{i-1,s/2} - u_s$. (Note that s is even.) Update also u_j for all $j > s$ according to (2) and $u_{2r-1} = w_{i-1,r} - u_{2r}$. If the test is failed and $s > 2|J|$, then do nothing. Update s to be $s - 2$ for both cases. This step including all updating is to be called a correction step of u_s , or simply a correction step, even for $s > |J|$. It does not necessarily mean that u_s was not $w(A_{i,s})$ just before the correction step though.

If $s \leq 2|J| + 8i^2 \log q$, repeat (iii) with updated s . Otherwise, let $w_{i,j} = u_j$, $j = 1, \dots, 2|J|$. Then return to the original label and update i , J to be $i + 1$, $\{j : w_{i,j} \text{ is defined and } |w_{i,j}| \geq \alpha\}$, respectively, and go to (ii).

(iv) Set $w_{i,j} = u_j$, $j = 1, \dots, 2|J|$. Then return to the original label and update J to be $\{j : w_{i,j} \text{ is defined and } |w_{i,j}| \geq \alpha\}$. If $i < \lceil 3 \log n \rceil$, then go to (ii) after updating i to be $i + 1$. If $i = \lceil 3 \log n \rceil$, then output J and declare that all coins in $\cup_{j \in J} A_{i,j}$ are counterfeit. Remove all coins that are declared counterfeit from the set A of all coins and update q to be $5q/6$. If $q > m^{0.8} + 2\epsilon m$ go to (i). Otherwise, go to (v).

(v) Apply the randomized binary search to find counterfeit coins one by one, using $(\lceil \log n \rceil + 3)(m^{0.8} + 2\epsilon m)$ queries.

The core parts of the algorithm are (ii) and (iii). If $w_{i-1,j} = w(A_{i-1,j})$ and every set $A_{i-1,j}$ contains at most one coin, then $w(A_{i,2j}) = 0$ or $w_{i-1,j}$. Provided $|\sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}}|$ is small enough, say less than $\alpha/2$ (see (a) lemma 3.3), it is not hard to show that $u_{2r} = w(A_{i,2r})$ and $u_{2r-1} = w(A_{i,2r-1})$ for all r . (See Corollary 3.5.) This was one of main ideas of Bshouty and Mazzawi [9]. In general, as some sets $A_{i-1,j}$ contain more than one counterfeit coin, u_{2r} may or may not be $w(A_{i,2r})$.

If r is the smallest r with $u_{2r} \neq w(A_{i,2r})$, $u_{2r'} = w(A_{i,2r'})$, $r' > r$, is not guaranteed any more even if the set $A_{i-1,r'}$ contains only one counterfeit coin. This is why we introduced random tests and correction steps in (iii). The random tests generate a random walk that travels according to the value of s . It turns out that the walk goes forward until it passes or at $2r$. Once the random walk passes or at $2r$, it goes backward with a probability close to 1 (not extremely close to 1 though). It is expected that the random walk with correction steps quickly identifies and corrects u_{2r} .

Moreover, it turns out that r is the smallest r with $u_{2r} \neq w(A_{i,2r})$ only if $A_{i-1,r}$ contains more than one counterfeit coin. If not many sets $A_{i-1,r}$ contain more than one counterfeit coin (see (c) of Lemma 3.2), the number of queries asked to identify and correct corresponding u_{2r} 's seems to be reasonably small. In other words, the lesser the number of sets $A_{i-1,r}$ containing more than one counterfeit coin is, the faster s increases. Eventually, s keeps increasing after all corresponding u_{2r} 's are corrected.

Remark. (a) Though the initial value -2 of s looks somewhat strange, it is natural as $s = 2r - 2$ when u_{2r} is corrected and the initial value must be determined as if the imaginary u_0 were corrected.

(b) When the random test fails, it may be tempting to find $A_{i,2r}$ with $w(A_{i,2r}) \neq 0$, say using a binary search. However, the number of queries needed to find such a set can be as large as $\Omega(\log q)$, while our algorithm is expected to correct u_{2r} using $O(i^2 \log(i^2 + 1))$ queries. This save queries when i is small. Though the bound is not extremely good if i is large, it is not really a matter as there are much less sets $A_{i-1,j}$ containing more than one counterfeit coin. (See (c) of Lemma 3.2.)

To analyze the algorithm, we precisely summarize core properties of the algorithm.

Lemma 3.3. Suppose (a)-(e) of Lemma 3.2 hold for q and $w_{i-1,j} = w(A_{i-1,j})$ for a fixed $i = 1, \dots, \lceil 3 \log n \rceil$ and all $j = 1, \dots, |J|$. Then we have the followings.

(a) For all $r = 1, \dots, |J|$, $\left| \sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}} \right| < \frac{\alpha}{2}$.

(b) If r is the smallest r such that $u_{2r} \neq w(A_{i,2r})$ when u_{2r} is first defined or updated, then neither $w(A_{i,2r-1})$ nor $w(A_{i,2r})$ is zero, especially $A_{i-1,r}$ contains more than one counterfeit coin.

(c) Suppose $i < \lceil 2 \log q \rceil$ and $u_j = w(A_{i,j})$ for all $j \leq 2r - 2$ at a step. If $s \leq 2r - 2$ at the step, then s keeps increasing until $s \geq 2r$. And once $s \geq 2r$, $s \geq 2r$ at all the following steps except possibly one step, which is a correction step of u_{2r} and $s = 2r - 2$.

Proof. (a) For $i < \lceil 2 \log q \rceil$, since $i \leq 2 \log q$, $2^{\gamma_i} \geq \frac{3\beta(i+2\log q)}{i\alpha}$ and $|w(A_{i,2(r+k)})| \leq \frac{\beta(i+2\log q)}{i}$ (as $A_{i,2(r+k)}$ contains at most $\frac{i+2\log q}{i}$ counterfeit coins by (b) of Lemma 3.2), we have

$$\left| \sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}} \right| \leq \sum_{k=1}^{k_r} \frac{\beta(i+2\log q)}{i \left(\frac{3\beta(i+2\log q)}{i\alpha} \right)^k} \leq \frac{\alpha}{3} + \frac{\alpha}{3} \sum_{k=1}^{\infty} \left(\frac{2\alpha \log q}{12\beta \log q} \right)^k \leq \frac{\alpha}{3} + \frac{\alpha}{3} \sum_{k=1}^{\infty} \left(\frac{1}{6} \right)^k < \frac{\alpha}{2}.$$

If $i \geq \lceil 2 \log q \rceil$, then each $A_{i,j}$ contains one or less counterfeit coin by (d) of Lemma 3.2, which together with $2^{\gamma_i} \geq \frac{6\beta}{\alpha}$ gives

$$\left| \sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}} \right| \leq \sum_{k=1}^{k_r} \frac{\beta}{\left(\frac{6\beta}{\alpha} \right)^k} \leq \frac{\alpha}{6} + \frac{\alpha}{6} \sum_{k=1}^{\infty} \left(\frac{\alpha}{6\beta} \right)^k \leq \frac{\alpha}{6} + \frac{\alpha}{6} \sum_{k=1}^{\infty} \left(\frac{1}{6} \right)^k < \frac{\alpha}{2}.$$

(b) As r is the smallest r such that $u_{2r} \neq w(A_{i,2r})$ when u_{2r} is defined or updated, $u_{2j} = w(A_{i,2j})$ for all $j < r$ and hence

$$w(A_{i,2r}) = x_r - \sum_{k=1}^{r-1} a_{rk} w(A_{i,2k}) - \sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}} = x_r - \sum_{k=1}^{r-1} a_{rk} u_{2k} - \sum_{k=1}^{k_r} \frac{w(A_{i,2(r+k)})}{2^{k\gamma_i}}.$$

If $u_{2r} = 0$, $u_{2r} \neq w(A_{i,2r})$ yields that $w(A_{i,2r}) \neq 0$. On the other hand, $u_{2r} = 0$ implies that $|x_r - \sum_{k=1}^{r-1} a_{rk} u_{2k}| < \alpha/2$. This together with (a) gives that $|w(A_{i,2r})| < \alpha$. Since $|w(A_{i-1,r})| = |w_{i-1,j}| \geq \alpha$ and $w(A_{i,2r-1}) = w(A_{i-1,r}) - w(A_{i,2r})$, $w(A_{i,2r-1}) \neq 0$. If $u_{2r} = w_{i-1,r} (= w(A_{i-1,r}))$, then $u_{2r} \neq w(A_{i,2r})$ yields that $w(A_{i,2r}) \neq w(A_{i-1,r})$ and hence $w(A_{i,2r-1}) = w(A_{i-1,r}) - w(A_{i,2r}) \neq 0$. On the other hand, $u_{2r} = w_{i-1,r}$ implies that $|x_r - \sum_{k=1}^{r-1} a_{rk} u_{2k}| \geq \alpha/2$. This together with (a) gives that $|w(A_{i,2r})| > 0$, i.e., $w(A_{i,2r}) \neq 0$.

(c) We prove this by reverse induction. For $r = |J| + 1$, if $u_j = w(A_{i,j})$ for all $j \leq 2r - 2 = 2|J|$, then $w(A_{i,j}) = 0$ whenever $u_j = 0$, for all $j \leq 2|J|$. Thus, the random test must be passed and s keeps increasing regardless of the value of s (as no u_j is updated). Suppose $u_j = w(A_{i,j})$ for all $j \leq 2r - 2$ with $r \leq |J|$. Then $w(A_{i,j}) = 0$ for all $j \leq 2r - 2$ with $u_j = 0$. If $s \leq 2r - 2$, the random test must be passed and hence s increases. Once $s > 2r - 2$, or equivalently $s \geq 2r$ (as s is even), no u_j with $j \leq 2r - 2$ is updated before a correction step of u_{2r} . If there is no correction step of u_{2r} , then $s \geq 2r$ at all the following steps. If u_{2r} is corrected at a step, then $s = 2r - 2$ and $u_j = w(A_{i,j})$ for all $j \leq 2r$ at the step. The induction hypothesis especially yields $s \geq 2r$ at all steps after the correction step.

□

To analyze (iii) of the algorithm for a fixed $i < \lceil 2 \log q \rceil$, we may regard the whole process as a random walk \mathcal{S} that travels according to the value of s . That is, $\mathcal{S} = (s_0, s_1, \dots)$, where s_k is the value of s at the end of the k^{th} step. Note that \mathcal{S} goes backward, i.e., s decreases, at a step if and only if the step is a correction step. We will see that \mathcal{S} goes forward until it passes or at $2r$ for the the smallest r with $u_{2r} \neq w(A_{i,2r})$, and then \mathcal{S} tends to go backward with probabilities close (not extremely though) to 1 until u_{2r} is corrected.

We partition \mathcal{S} into a few subrandom walks that are essentially independent identically distributed (i.i.d). They are not exactly i.i.d though. Let $r_0 = 0$. The 0^{th} (sub)random walk \mathcal{S}_0 (of \mathcal{S}) starts when the whole process starts and ends at the same time, that is, $\mathcal{S}_0 = (s_0)$ (recall $s_0 = -2$). Let r_1 be the the smallest r with $u_{2r} \neq w(A_{i,2r})$. The first random walk \mathcal{S}_1 starts immediately after \mathcal{S}_0 ends and it ends when $s = 2r_1 - 2$ at a backward step or $s > 2|J| + 8i^2 \log q$ for the first time, whichever comes first. Generally, for $\ell \geq 2$, if $\mathcal{S}_{\ell-1}$ ends with $s = 2r_{\ell-1} - 2$, then let r_ℓ be the smallest $r \leq |J|$ such that $u_{2r} \neq w(A_{i,2r})$ at the step $\mathcal{S}_{\ell-1}$ ends. The ℓ^{th} random walk \mathcal{S}_ℓ starts immediately after $\mathcal{S}_{\ell-1}$ ends, and it ends when $s = 2r_\ell - 2$

at a backward step or $s > 2|J| + 8i^2 \log q$ for the first time, whichever comes first. However, \mathcal{S}_ℓ does not end at a forward step with $s = 2r_\ell - 2$. In theory, it is possible that \mathcal{S}_ℓ is infinite, though it is not difficult to show that the probability of \mathcal{S}_ℓ being infinite is 0. Both of $r_{\ell'}$ and $\mathcal{S}_{\ell'}$ are not defined for all $\ell' \geq \ell$, if $\mathcal{S}_{\ell-1}$ is infinite or ends with $s > 2|J| + 8i^2 \log q$, or $u_{2r} = w(A_{i,2r})$ for all $r \leq |J|$ at the last step of $\mathcal{S}_{\ell-1}$.

The random walk \mathcal{S}_ℓ is called good if it is defined and ends with $s = 2r_\ell - 2$. Note that the last step of a good random walk \mathcal{S}_ℓ is the first correction step of u_{2r_ℓ} after \mathcal{S}_ℓ starts. In other words, a good random walk \mathcal{S}_ℓ ends when it corrects u_{2r_ℓ} where r_ℓ is the smallest r such that $u_{2r} \neq w(A_{i,2r})$ when it starts. We also note that r_ℓ , \mathcal{S}_ℓ are defined only if $\mathcal{S}_{\ell-1}$ is good. In particular, \mathcal{S}_ℓ is good only if $\mathcal{S}_{\ell'}$ is good for all $\ell' \leq \ell - 1$.

Corollary 3.4. *Under the same hypotheses as in Lemma 3.3 with $i \leq \lceil 2 \log q \rceil - 1$, we have the followings.*

- (a) *For $\ell \geq 1$, if \mathcal{S}_ℓ is good, $u_j = w(A_{i,j})$ for all $1 \leq j \leq 2r_\ell$ at the last step of \mathcal{S}_ℓ , especially $r_{\ell+1} > r_\ell$ if $r_{\ell+1}$ is defined. Furthermore, a good random walk \mathcal{S}_ℓ starts with $s = 2r_{\ell-1} - 2$ and keeps going forward until $s \geq 2r_\ell$, and then goes back and force with $s \geq 2r_\ell$ at all steps except the last step at which $s = 2r_\ell - 2$.*
- (b) *If r_ℓ is defined, then neither $w(A_{i,2r_\ell-1})$ nor $w(A_{i,2r_\ell})$ is zero and $A_{i-1,r}$ contains more than one counterfeit coin. In particular, r_ℓ and \mathcal{S}_ℓ are not defined if $\ell > h_q := \lceil 2^{-(i+1)}q + q^{3/4} \rceil$.*
- (c) *Suppose every \mathcal{S}_ℓ is good if defined. Then $w_{i,j} = w(A_{i,j})$, for all $j = 1, \dots, 2|J|$.*

Proof. (a) For $\ell \geq 1$, suppose $u_j = w(A_{i,j})$ for all $j \leq 2r_{\ell-1}$ at the last step of $\mathcal{S}_{\ell-1}$. (This is trivial when $\ell = 1$.) Since \mathcal{S}_ℓ is good only if $\mathcal{S}_{\ell-1}$ is good, the induction hypothesis may be applied to obtain $r_\ell > r_{\ell-1}$ and hence

$$u_j = w(A_{i,j}) \quad \text{for all } j \leq 2r_\ell - 2$$

at the last step of $\mathcal{S}_{\ell-1}$. Then, (c) of Lemma 3.3 gives that s keeps increasing (without updating u_j) after the last step of $\mathcal{S}_{\ell-1}$, at which $s = 2r_{\ell-1} - 2$, until $s \geq 2r_\ell$. Once $s \geq 2r_\ell$, no u_j with $j \leq 2r_\ell - 2$ is updated before the last step of \mathcal{S}_ℓ . Since \mathcal{S}_ℓ is good, u_{2r} is corrected and hence $u_{2r} = w(A_{i,2r})$, $u_{2r-1} = w(A_{i,2r-1})$ at the last step of \mathcal{S}_ℓ . The second part is already shown too.

(b) Since $r_\ell > r_{\ell-1}$ is defined, $s = r_{\ell-1} - 2$ at the last step of $\mathcal{S}_{\ell-1}$ and u_{2r_ℓ} is updated at the last step of $\mathcal{S}_{\ell-1}$. By (b) of Lemma 3.3, neither $w(A_{i,2r-1})$ nor $w(A_{i,2r})$ is zero. The second part follows from that all r_ℓ 's are distinct and there are at most h_q sets $A_{i-1,r}$ containing more than one counterfeit coin (see (c) of Lemma 3.2).

(c) For the largest ℓ for which r_ℓ is defined, as \mathcal{S}_ℓ is good and $r_{\ell+1}$ is not defined, $u_{2r} = w(A_{i,2r})$ for all $r = 1, \dots, |J|$ at the last step of \mathcal{S}_ℓ . Since s keeps increasing after the last step and eventually $s > 2|J| + 8i^2 \log q$ without updating u_j 's, we have $w_{i,2r} = w(A_{i,2r})$ for $r = 1, \dots, |J|$, and $w_{i,2r-1} = w_{i-1,r} - w_{i,2r} = w(A_{i-1,r}) - w(A_{i,2r}) = w(A_{i,2r-1})$.

□

If $\lceil 2 \log q \rceil \leq i \leq \lceil 3 \log n \rceil$, each $A_{i,j}$ contains at most one counterfeit coin by (d) of Lemma 3.2. Then it easily follows that $w_{i,j} = w(A_{i,j})$ for all $j = 1, \dots, 2|J|$.

Corollary 3.5. *Under the same hypotheses as in Lemma 3.3 with $\lceil 2 \log q \rceil \leq i \leq \lceil 3 \log n \rceil$, $w_{i,j} = w(A_{i,j})$ for all $j = 1, \dots, 2|J|$.*

Proof. Recall that $w_{i,j} = u_j$, where u_j 's are defined in (ii). Take, if any, the smallest r such that $u_{2r} \neq w(A_{i,2r})$. Then, (b) of Lemma 3.3 implies that $A_{i-1,r}$ contains more than one counterfeit coin, which is not possible as each $A_{i-1,r}$ contains at most one counterfeit coin due to (d) of Lemma 3.2. Hence, $w_{i,2r} = w(A_{i,2r})$ for all $r = 1, \dots, |J|$ and $w_{i,2r-1} = u_{2r-1} = w(A_{i-1,r}) - w(A_{i,2r}) = w(A_{i,2r-1})$.

□

Corollaries 3.4 and 3.5 provide all but one basic properties to analyze the algorithm. The missing property is that, with high probability, every \mathcal{S}_ℓ is good if defined, the hypothesis of (c) of Corollary 3.4.

For the query complexity, an upper bound for the number $|\mathcal{S}_\ell|$ of steps in \mathcal{S}_ℓ is needed. As we want to bound $|\mathcal{S}_\ell|$ only for good \mathcal{S}_ℓ , we will consider $|\mathcal{S}_\ell|\chi_\ell$, where

$$\chi_\ell = \begin{cases} 1 & \text{if } \mathcal{S}_\ell \text{ is good} \\ 0 & \text{otherwise.} \end{cases}$$

It will be first shown that, after \mathcal{S}_ℓ passes or at $2r_\ell$, the random walk goes to backward with probability at least $1 - \frac{1}{8(i^2+1)}$ until u_{2r_ℓ} is corrected, which follows from $w(\cup\{A_{i,j} : \text{selected } j\}) \neq 0$ with probability at least $1/2$ during the process. Thus, \mathcal{S}_ℓ goes backward by at least $7/4$ in expectation after \mathcal{S}_ℓ passes or at $2r_\ell$, as \mathcal{S}_ℓ goes forward by $2i^2$ with probability at most $\frac{1}{8(i^2+1)}$ and goes backward by 2 otherwise. This is why \mathcal{S}_ℓ is expected to be good. The number F_ℓ of forward steps in \mathcal{S}_ℓ after it passes or at $2r_\ell$ is also expected to be reasonably small, namely $O(1)$ with a probability close to 1. It actually turns out that the probability of $F_\ell = k$ is at most e^{-k+1} and the sum $\sum_{\ell=1}^{h_q} F_\ell$ may be bounded by $O(h_q)$ with high enough probability, say with probability $1 - e^{-\Omega(q^{3/4})}$, where $h_q = \lceil 2^{-(i+1)}q + q^{3/4} \rceil$ as in (b) of Lemma 3.4.

Then, it is not difficult to show that the number of all steps in \mathcal{S}_ℓ after it passes or at $2r_\ell$ is $O(i^2 F_\ell)$, especially there are $O(i^2 F_\ell)$ backward steps in \mathcal{S}_ℓ by (a) of Corollary 3.4. Therefore, there are $O(i^2 h_q)$ backward steps in \mathcal{S} with high probability. All other steps in \mathcal{S} are forward steps and hence there are

$$O\left(i^2 h_q + \frac{|J| + 8i^2 \log q + 2i^2 h_q}{2i^2}\right) = O\left(\frac{|J|}{i^2} + (i^2 + 2)\left(\frac{q}{2^{i+1}} + q^{3/4}\right)\right)$$

steps in \mathcal{S} . As $O(\log(i^2 + 1))$ queries are asked at each step, the number of queries asked in the i^{th} round, $i = 1, \dots, \lceil 2 \log q \rceil - 1$, is $O\left(q\left(\frac{1}{i^2} + \frac{i^2}{2^i}\right) \log(i^2 + 1)\right)$ assuming $|J| \leq q$.

The precise statements are presented in the next lemma. Though idea is simple as illustrated above, our proof of the lemma is somewhat lengthy, partly because it is proven rigorously without referring other theories. We prove the lemma at the end of this section. Readers familiar with random walks may skip the proof.

Lemma 3.6. *Under the same hypotheses as in Lemma 3.3 with $i \leq \lceil 2 \log q \rceil - 1$, if $u_{2r} \neq w(A_{i,2r})$ and $s \geq 2r$ at a step, then the probability that s increases at the next step is at most $\frac{1}{8(i^2+1)}$. Moreover, every \mathcal{S}_ℓ is good if defined, with probability $1 - O(q^{-3})$, and the number $|\mathcal{S}|$ of all steps satisfies*

$$\Pr \left[|\mathcal{S}| \geq \frac{|J|}{i^2} + 4(i^2 + 2)\left(\frac{q}{2^{i+1}} + q^{3/4}\right) \right] = O(q^{-3}).$$

Correctness of the algorithm Once Lemmas 3.2, 3.6 and Corollaries 3.4, 3.5 are established, it is easy to see that the algorithm finds counterfeit coins as desired. In the next lemma, we precisely describe it along with a property needed to bound query complexity.

Lemma 3.7. *For a fixed $q > m^{0.8} + 2\epsilon m$, the followings hold with probability $1 - O(1/q)$, assuming the same in the prior round.*

- (a) *The statements (a)-(e) of Lemma 3.2 hold.*
- (b) *Whenever $w_{i,j}$ is defined, $w_{i,j} = w(A_{i,j})$. In particular, a coin declared to be counterfeit must be counterfeit.*
- (c) *The algorithm finds every counterfeit coin c with $|w(c)| \geq \alpha$ that is a unique counterfeit coin of $A_{0,\ell}$ for some $\ell = 1, \dots, 2^{\ell_q}$. And the number of remaining counterfeit coins is at most the updated q .*
- (d) *The number of queries asked in all rounds of (iii) is $O(q)$, where the constant in $O(q)$ is at most $\sum_{i=1}^{\infty} \frac{5+\log(i^2+1)}{i^2} + \frac{(i^2+2)(5+\log(i^2+1))}{2^{i-1}} + o(1)$.*

Proof. As $q > 2\epsilon m$ and (c) holds in the prior round, Lemma 3.2 yields that the statements in (a) hold with probability $1 - O(1/q)$. We assume the statements to prove the other properties.

To prove the other properties, we further assume that every \mathcal{S}_ℓ is good if defined and that, for each $i = 1, \dots, \lceil 2 \log q \rceil - 1$ and the number $|\mathcal{S}|$ of all steps in the i^{th} round of (iii),

$$|\mathcal{S}| \leq \frac{|J|}{i^2} + 4(i^2 + 2) \left(\frac{q}{2^{i+1}} + q^{3/4} \right), \quad (3)$$

both of which hold with probability $1 - O(q^{-3})$ by Lemma 3.6. Then the first part of (b) follows from (c) of Corollary 3.4, and Corollary 3.5. Since every $A_{\lceil 3 \log n \rceil, j}$ consists of one or no coin, each coin c in $\cup_{j \in J} A_{\lceil 3 \log n \rceil, j}$ satisfies $|w(c)| = |w(A_{\lceil 3 \log n \rceil, j})| = |w_{\lceil 3 \log n \rceil, j}| \geq \alpha$ for some $j \in J$, especially, c is counterfeit.

If a coin c with $|w(c)| \geq \alpha$ is a unique counterfeit coin in A_{0, ℓ_0} , then, for each $i = 0, \dots, \lceil 3 \log n \rceil$, there is a unique ℓ_i such that $A_{i, \ell_i} \subseteq A_{0, \ell_0}$ contains c . It is clear, by the way how $A_{i, j}$'s are constructed, that $A_{i, \ell_i} \subseteq A_{i-1, \ell_{i-1}}$ for all $i = 1, \dots, \lceil 3 \log n \rceil$. Moreover, since c is a unique counterfeit coin of A_{i, ℓ_i} , $|w(A_{i, \ell_i})| \geq \alpha$ for all $i = 0, \dots, \lceil 3 \log n \rceil$. For $i = 0$, $\ell_0 \in J$ as $|w_{0, \ell_0}| = |w(A_{0, \ell_0})| \geq \alpha$. For $i \geq 1$, assuming $\ell_{i-1} \in J$ when the prior round ends, $w_{i, \ell_i} = w(A_{i, \ell_i})$ by (c) of Corollary 3.4, as $\ell_{i-1} \in J$ and $A_{i, \ell_i} \subseteq A_{i-1, \ell_{i-1}}$. Thus, $|w_{i, \ell_i}| = |w(A_{i, \ell_i})| \geq \alpha$ implies that ℓ_i is in the updated J . We have just shown that $\ell_i \in J$ when the i^{th} round ends for each i , especially, for $i = \lceil 3 \log n \rceil$. As $c \in A_{i, \ell_i}$ for $i = \lceil 3 \log n \rceil$, c is declared to be counterfeit. The second part of (c) follows from (a) of Lemma 3.2.

Note that $|J| = |\{j : w_{i, j} \text{ is defined and } |w_{i, j}| \geq \alpha\}| \leq q$ by the second part of (c) in the prior round and first part of (b), as $|w_{i, j}| = |w(A_{i, j})| \geq \alpha$ implies that $A_{i, j}$ contains a counterfeit coin and the number of such sets is at most the number of counterfeit coins. Since the algorithm asks at most $5 + \log(i^2 + 1)$ queries at each step of \mathcal{S} (one more query is needed in backward steps), (3) yields that the number of queries is at most

$$\sum_{i=1}^{\lceil 2 \log q \rceil - 1} \left(\frac{(5 + \log(i^2 + 1))q}{i^2} + 4(i^2 + 2)(5 + \log(i^2 + 1)) \left(\frac{q}{2^{i+1}} + q^{3/4} \right) \right) = O(q).$$

□

The lemma especially says that the number of remaining counterfeit coins decreases by factor $5/6$, with probability $1 - O(1/q)$. Applying this inductively until $q \leq m^{0.8} + 2\epsilon m$, we know the algorithm find all but at most $m^{0.8} + 2\epsilon m$ counterfeit coins before it goes to (v), with probability $1 - O(1/m^{0.8})$. All the remaining $m^{0.8} + 2\epsilon m$ counterfeit coins are found in (v), with probability $1 - e^{-\Omega(m^{0.8})}$, by Lemma 3.1.

Corollary 3.8. *The algorithm find all but at most $m^{0.8} + 2\epsilon m$ counterfeit coins before it goes to (v), with probability $1 - O(1/m^{0.8})$. All the remaining $m^{0.8} + 2\epsilon m$ counterfeit coins are found in (v), with probability $1 - e^{-\Omega(m^{0.8})}$, by Lemma 3.1.*

Query Complexity Suppose (a)-(d) of Lemma 3.7 hold for all q , which occurs with probability $1 - O(1/m^{0.8})$. Then for each q , the number of remaining counterfeit coins is at most q . Especially, $|J| \leq q$ as seen in last paragraph of the proof of Lemma 3.7. For each q , $2^{\ell_q} \leq 2q$ queries are needed in (i). For each q and i , the number of queries asked in (ii) is

$$\frac{(2 + o(1))\gamma_i |J|}{\log(\gamma_i |J|)} \leq \begin{cases} \frac{(2+o(1))|J|}{\log |J|} \left\lceil \log\left(\frac{3\beta(i+2\log q)}{i\alpha}\right) \right\rceil & \text{if } i < \lceil 2 \log q \rceil \\ \frac{(2+o(1))|J|}{\log |J|} \left\lceil \log(6\beta/\alpha) \right\rceil & \text{if } i \geq \lceil 2 \log q \rceil. \end{cases}$$

Since $|J| \leq q$ and

$$\sum_{i=1}^{\lceil 2 \log q \rceil - 1} \left\lceil \log\left(\frac{3\beta(i+2\log q)}{i\alpha}\right) \right\rceil \leq 4 \log q \log(3\beta/\alpha) + \log \left(\frac{2\lceil 2 \log q \rceil - 1}{\lceil 2 \log q \rceil - 1} \right) \leq 4 \log q \log(3\beta/\alpha) + 4 \log q + 1,$$

and

$$\sum_{i=\lceil 2 \log q \rceil}^{\lceil 3 \log n \rceil} \lceil \log(6\beta/\alpha) \rceil \leq 3 \log(6\beta/\alpha) \log n + 3 \log n$$

for each q , the number of queries asked in (ii) is $O\left(\frac{q \log(\beta/\alpha) \log n}{\log q}\right)$.

The number of all queries asked in (iii) for each q is $O(q)$ by (d) of Lemma 3.7. No query is asked in (iv) and hence the total number of queries asked for fixed $q > m^{0.8} + 2\epsilon m$ is $O\left(\frac{q \log(\beta/\alpha) \log n}{\log q}\right)$. As q keeps decreasing by factor of $5/6$, the number of queries asked before the algorithm goes to (v) is $O\left(\frac{m \log(\beta/\alpha) \log n}{\log m}\right)$. \square

This together with Corollary 3.8 implies that, if we artificially stop the algorithm when it asks $\frac{\eta m \log(\beta/\alpha) \log n}{\log m}$ queries, for the constant η in the $O\left(\frac{m \log(\beta/\alpha) \log n}{\log m}\right)$ term, all but at most $m^{0.8} + 2\epsilon m$ counterfeit coins are found with probability $1 - O(1/m^{0.8})$. As $(\lceil \log n \rceil + 3)(m^{0.8} + 2\epsilon m)$ queries are asked in (v) of the algorithm, Theorem 1.2 follows. We conclude this section by proving Lemma 3.6.

Proof of Lemma 3.6 Note that each u_{2r} may have one of three values $0, w_{i-1,r}, w(A_{i,2r})$. Since $u_{2r} \neq w(A_{i,2r})$, u_{2r} is either 0 or $w_{i-1,j}$. If $u_{2r} = 0$, then $w(A_{i,2r}) \neq 0$. If $u_{2r} = w_{i-1,j} (= w(A_{i-1,j}))$, then $u_{2r-1} = 0$ while $w(A_{i,2r}) \neq u_{2r} = w(A_{i-1,j})$ yields $w(A_{i,2r-1}) = w(A_{i-1,j}) - w(A_{i,2r}) \neq 0$. Particularly, there is $\ell \leq s$ such that $w(A_{i,\ell}) \neq 0$ while $u_\ell = 0$. Suppose the random selection other than ℓ is carried out. Then the set of coins to be weighed is either $\cup\{A_{i,j} : \text{selected } j, j \neq \ell\}$ or $\cup\{A_{i,j} : \text{selected } j, j \neq \ell\} \cup A_{i,\ell}$, each of which occurs with probability $1/2$. Since $w(A_{i,\ell}) \neq 0$ implies that the weights of the two sets are different,

$$\Pr[w(\cup\{A_{i,j} : \text{selected } j\}) = 0] \leq 1/2.$$

After independently performing this $\lceil \log(i^2 + 1) \rceil + 3$ times, the probability that all weights are 0 is at most $2^{\lceil \log(i^2 + 1) \rceil + 3} \leq \frac{1}{8(i^2 + 1)}$. That is, s increases at the next step with probability at most $\frac{1}{8(i^2 + 1)}$.

For the second part, suppose \mathcal{S}_ℓ is defined but it is not good, which especially means that $\mathcal{S}_{\ell-1}$ is good. Then \mathcal{S}_ℓ must be infinite or reach a step with $s > 2|J| + 8i^2 \log q$. As \mathcal{S}_ℓ starts with $s = 2r_{\ell-1} - 2$, $r_{\ell-1} < r_\ell$, and $u_{2r} = w(A_{i,2r})$ for all $r \leq 2r_\ell - 2$, the random walk \mathcal{S}_ℓ keeps going forward until $s \geq 2r_\ell$ by (c) of Lemma 3.3. Let σ_ℓ be the value of s when \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time. Then

$$2r_\ell \leq \sigma_\ell \leq 2r_\ell + 2i^2 - 2, \quad \text{or} \quad 0 \leq \sigma_\ell/2 - r_\ell \leq i^2 - 1, \quad (4)$$

for s increases by $2i^2$. Hence, there must be at least $\lfloor 4 \log q \rfloor$ more forwarding steps to reach a step with $s > 2|J| + 8i^2 \log q$, as, otherwise,

$$s \leq \sigma_\ell + 2i^2(\lfloor 4 \log q \rfloor - 1) \leq 2r_\ell - 2 + 2i^2 + 2i^2(\lfloor 4 \log q \rfloor - 1) \leq 2|J| + 8i^2 \log q.$$

If \mathcal{S}_ℓ is infinite, there must be at least $\lfloor 4 \log q \rfloor$ more forwarding steps too.

Counting after \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time, let T be the number of steps in \mathcal{S}_ℓ until there are $\lfloor 4 \log q \rfloor$ more forwarding steps. For \mathcal{S}_ℓ is not good, there is no correction step of u_{2r_ℓ} , or equivalently $s \geq 2r_\ell$ after the count starts, particularly, T satisfies

$$\sigma_\ell + 2i^2 \lfloor 4 \log q \rfloor - 2(T - \lfloor 4 \log q \rfloor) \geq 2r_\ell,$$

which, together with (4), gives

$$T \leq (i^2 + 1) \lfloor 4 \log q \rfloor + \sigma_\ell/2 - r_\ell \leq (i^2 + 1)(\lfloor 4 \log q \rfloor + 1).$$

We have just shown that, for $t = (i^2 + 1)(\lfloor 4 \log q \rfloor + 1)$,

$$\Pr[\mathcal{S}_\ell \text{ is not good}] \leq \Pr \left[\exists \lfloor 4 \log q \rfloor \text{ forward steps among the first } t \text{ or less steps of } \mathcal{S}_\ell \right]. \quad (5)$$

To bound the last probability, it is convenient to introduce an auxiliary random walk \mathcal{S}_ℓ^* . The infinite random walk \mathcal{S}_ℓ^* starts when \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time and it is the same as \mathcal{S}_ℓ until \mathcal{S}_ℓ ends. Once \mathcal{S}_ℓ ends, \mathcal{S}_ℓ^* keeps going forward by $2i^2$ with probability $\frac{1}{8(i^2+1)}$ and backward by 2 with probability $1 - \frac{1}{8(i^2+1)}$. Then, at any step, \mathcal{S}_ℓ^* goes forward with probability at most $\frac{1}{8(i^2+1)}$.

As there are $\lfloor 4 \log q \rfloor$ forward steps among the first t steps of \mathcal{S}_ℓ^* if there are $\lfloor 4 \log q \rfloor$ forward steps among the first t or less steps of \mathcal{S}_ℓ , (5) gives

$$\Pr[\mathcal{S}_\ell \text{ is not good}] \leq \Pr \left[\exists \lfloor 4 \log q \rfloor \text{ forward steps among the first } t \text{ steps of } \mathcal{S}_\ell^* \right],$$

which is at most $\binom{t}{\lfloor 4 \log q \rfloor} \left(\frac{1}{8(i^2+1)} \right)^{\lfloor 4 \log q \rfloor}$. Therefore, using $\binom{t}{k} \leq \left(\frac{et}{k} \right)^k$,

$$\Pr[\mathcal{S}_\ell \text{ is not good}] \leq \binom{t}{\lfloor 4 \log q \rfloor} \left(\frac{1}{8(i^2+1)} \right)^{\lfloor 4 \log q \rfloor} \leq \exp \left(\lfloor 4 \log q \rfloor \ln \frac{e(i^2+1)(\lfloor 4 \log q \rfloor + 1)}{8(i^2+1)\lfloor 4 \log q \rfloor} \right).$$

Using $\ln(e/8) \leq -1$ and $\ln(1+y) \leq y$ for $y \geq 0$, we obtain

$$\Pr[\mathcal{S}_\ell \text{ is not good}] \leq \exp \left(-\lfloor 4 \log q \rfloor + 1 \right) = O(q^{-4}).$$

Since \mathcal{S}_ℓ is defined for at most h_q indices ℓ by (b) of Corollary 3.4, and $h_q = O(q)$, Boole's inequality yields the desired bound.

For the last bound, if \mathcal{S}_ℓ is good, let F_ℓ be the number of all forward steps in \mathcal{S}_ℓ after \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time. If \mathcal{S}_ℓ is not good or not defined, then $F_\ell = 0$. If $F_\ell = k \geq 1$, then \mathcal{S}_ℓ must be good and, for the number t of all steps in \mathcal{S}_ℓ after \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time, we have

$$\sigma_\ell + 2i^2k - 2(t - k) = 2r_\ell - 2 \quad \text{or} \quad t = (i^2 + 1)k + \sigma_\ell/2 - r_\ell + 1 \leq (i^2 + 1)(k + 1),$$

(recall that σ_ℓ is the value of s when \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time).

After \mathcal{S}_ℓ reaches a step with $s \geq 2r_\ell$ for the first time, the probability that \mathcal{S}_ℓ moves forward is at most $\frac{1}{8(i^2+1)}$ until it ends. Moreover, the bound for the probability holds regardless of $F_{\ell'}$, $\ell' < \ell$. The same argument as above gives, for a positive integer k ,

$$\Pr[F_\ell = k | F_1, \dots, F_{\ell-1}] \leq \Pr \left[\exists k \text{ forward steps among the first } t \text{ steps of } \mathcal{S}_\ell^* \right]$$

and, by $\binom{t}{k} \leq \left(\frac{et}{k} \right)^k$, $\ln(e/8) \leq -1$ and $\ln(1+y) \leq y$ for $y > 0$,

$$\Pr[F_\ell = k | F_1, \dots, F_{\ell-1}] \leq \binom{t}{k} \left(\frac{1}{8(i^2+1)} \right)^k \leq \exp \left(k \ln \frac{et}{8k(i^2+1)} \right) \leq e^{-k+1}.$$

The inequality still holds when $k = 0$. For $h = h_q = \lfloor 2^{-(i+1)}q + q^{3/4} \rfloor$,

$$\Pr \left[F_1 = k_1, \dots, F_h = k_h \right] = \prod_{\ell=1}^h \Pr \left[F_\ell = k_\ell \mid F_1 = k_1, \dots, F_{\ell-1} = k_{\ell-1} \right] \leq e^{-(\sum_{\ell=1}^h k_\ell) + h},$$

implies that

$$\Pr \left[\sum_{\ell=1}^h F_\ell = k \right] = \sum_{\substack{k_\ell \geq 0 \\ k_1 + \dots + k_h = k}} \Pr \left[F_1 = k_1, \dots, F_h = k_h \right] \leq \binom{k+h}{h} e^{-k+h}.$$

Since $\binom{k+h}{h} \leq \left(\frac{e(k+h)}{h} \right)^h$, we have

$$\Pr \left[\sum_{\ell=1}^h F_\ell = k \right] \leq \exp \left(h \ln \frac{e(k+h)}{h} - k + h \right) = \exp \left(h \ln \frac{(k+h)}{h} - k + 2h \right).$$

For $k \geq 4h - 1$, $h \ln \frac{(k+h)}{h} - k + 2h \leq -4k/5 + 3h$ yields that

$$\Pr \left[\sum_{\ell=1}^h F_{\ell} \geq 4h - 1 \right] = \sum_{k=4h-1}^{\infty} \Pr \left[\sum_{\ell=1}^h F_{\ell} = k \right] \leq \sum_{k=4h-1}^{\infty} e^{-4k/5+3h} \leq 2e^{-(h-4)/5}.$$

Finally, for good \mathcal{S}_{ℓ} , the number of forward steps in \mathcal{S}_{ℓ} is

$$\frac{\sigma_{\ell} - (2r_{\ell-1} - 2)}{2i^2} + F_{\ell} = \frac{r_{\ell} - r_{\ell-1}}{i^2} + \frac{\sigma_{\ell}/2 - r_{\ell} + 1}{i^2} + F_{\ell},$$

while the number backward steps in \mathcal{S}_{ℓ} is,

$$\frac{1}{2} \left(\sigma_{\ell} - (2r_{\ell} - 2) + 2i^2 F_{\ell} \right) = \sigma_{\ell}/2 - r_{\ell} + 1 + i^2 F_{\ell}.$$

As $\sigma_{\ell}/2 - r_{\ell} \leq i^2 - 1$ by (4),

$$|\mathcal{S}_{\ell}|_{\chi_{\ell}} \leq \frac{r_{\ell} - r_{\ell-1}}{i^2} + 1 + F_{\ell} + i^2 + i^2 F_{\ell} = \frac{r_{\ell} - r_{\ell-1}}{i^2} + (i^2 + 1)(F_{\ell} + 1).$$

Therefore,

$$\Pr \left[\sum_{\ell=1}^h |\mathcal{S}_{\ell}|_{\chi_{\ell}} \geq \frac{r^*}{i^2} + 4(i^2 + 1)h \right] \leq \Pr \left[\sum_{\ell=1}^h F_{\ell} \geq 4h - 1 \right] \leq 2e^{-(h-4)/5} \leq 2e^{-q^{3/4}/5+1},$$

where $r^* = \max\{r_{\ell} : \mathcal{S}_{\ell} \text{ is good}\}$.

Suppose every \mathcal{S}_{ℓ} is good if defined. Then there are $\lceil \frac{2|J| - (2r^* - 2) + 8i^2 \log q}{2i^2} \rceil$ more steps after u_{2r^*} is corrected, and the number $|\mathcal{S}|$ of all steps in \mathcal{S} , or equivalently in (iii) for fixed i , is

$$\left\lceil \frac{|J| - r^* + 1 + 4i^2 \log q}{i^2} \right\rceil + \sum_{\ell=1}^h |\mathcal{S}_{\ell}|_{\chi_{\ell}} \leq \frac{|J| - r^* + 1}{i^2} + 4 \log q + 1 + \sum_{\ell=1}^h |\mathcal{S}_{\ell}|_{\chi_{\ell}}.$$

Thus, if $\sum_{\ell=1}^h |\mathcal{S}_{\ell}|_{\chi_{\ell}} < \frac{r^*}{i^2} + 4(i^2 + 1)h$, then

$$|\mathcal{S}| < \frac{|J|}{i^2} + 4(i^2 + 1)h + \frac{1}{i^2} + 4 \log q + 1 < \frac{|J|}{i^2} + 4(i^2 + 2) \left(\frac{q}{2^{i+1}} + q^{3/4} \right).$$

By the contrapositive, if $|\mathcal{S}| \geq \frac{|J|}{i^2} + 4(i^2 + 2) \left(\frac{q}{2^{i+1}} + q^{3/4} \right)$, then either there is \mathcal{S}_{ℓ} that is defined but not good or

$$\sum_{\ell=1}^h |\mathcal{S}_{\ell}|_{\chi_{\ell}} \geq \frac{r_{\ell}}{i^2} + 4(i^2 + 1)h,$$

which gives

$$\Pr \left[|\mathcal{S}| \geq \frac{|J|}{i^2} + 4(i^2 + 2) \left(\frac{q}{2^{i+1}} + q^{3/4} \right) \right] = O(q^{-3} + e^{-q^{3/4}/5}) = O(q^{-3}).$$

□

4 Finding Weighted Graphs

In this section, we present a randomized algorithm finding weighted graphs using additive queries, where an additive query asks the sum of weights of edges with both ends in a fixed set. The algorithm uses coin weighing algorithms presented in the previous section.

Let $G = (V, E, w_G)$ be a weighted graph with $w_G(e) \neq 0$ for all $e \in E$. We just say graphs for weighted graphs. First of all, it is enough to consider bipartite graphs: For general graphs, one may consider two disjoint copies X, Y of V . The copy of $u \in V$ in X and the copy of $v \in V$ in Y form an edge if and only if uv is an edge in G , and, of course, the weight is inherited. Then a query of type $w(A, B) := \sum_{x \in A, y \in B} w(x, y)$, $A \subset X, B \subset Y$ is a linear combination of four additive queries in G , that is,

$$w(A, B) = w_G(A \cup B) - w_G(A \setminus B) - w_G(B \setminus A) + w_G(A \cap B). \quad (6)$$

In the rest of this section, we consider weighted bipartite graphs $G = (X \cup Y, E, w)$ with $|X| = |Y| = n$ and $|E| \leq m$. A query means that one takes two sets $A \subset X$ and $B \subset Y$ and finds out $w(A, B) := \sum_{a \in A, b \in B} w(a, b)$.

If $\mathcal{O}(m \log n)$ queries are allowed, it is easy to find the graph using the randomized binary search:

Randomized Binary Search for Graph Suppose $n, m \geq 1$ and a bipartite graph $G = X \cup Y$ with at most m edges and $|X|, |Y| \leq n$ is given. Then, take random subsets X', Y' of X and Y , respectively, so that each vertex $x \in X$ ($y \in Y$, resp.) in X' (Y' , resp.) with probability $1/2$, independently of all other vertices. If $w(X', Y') \neq 0$, find an edge there using the deterministic binary search. Otherwise, take a new random sets X', Y' and do it again. Stop when $(2\lceil \log n \rceil + 5)m$ queries are asked. Output all edges found.

The deterministic binary search means that divide X' into two parts X'_1, X'_2 with size difference at most 1. If $w(X'_1, Y') \neq 0$ take X'_1 , otherwise, take X'_2 . Keep doing this until a vertex x with $w(x, Y') \neq 0$ is found. Then, find $y \in Y'$ with $w(x, y) \neq 0$ using the same method.

If there is an edge in G , the probability of $w(X', Y') \neq 0$ is at least $1/4$. It may be shown that $(2\lceil \log n \rceil + 4 + o(1))m$ queries are enough to find all edges in G , with high probability. We may prove $(2\lceil \log n \rceil + 5)m$ queries are enough with probability $1 - e^{-\Omega(m)}$, a proof of which is presented in Appendix.

Lemma 4.1. *The randomized binary search finds all edges of G with probability $1 - e^{-\Omega(m)}$.*

For a better query complexity, a more sophisticated algorithm is needed. We first present an algorithm finding all edges of G when the maximum degree of G is small, say at most $m^{0.1}$. Then another algorithm is introduced to find vertices of large degree and edges containing them. Concatenating two algorithms, the following theorem may be shown.

Theorem 4.2. *Let n, m be positive integers with $n^2 \geq m \geq 2$ and let $\alpha, \beta > 0$ be positive real numbers (not necessarily constants) with $2\alpha < \beta$. Suppose a bipartite (weighted) graph G is given such that each part of G has at most n vertices and there are m or less edges in G . If the weights $w(e)$ of edges satisfy $\alpha \leq |w(e)| \leq \beta$, then there is a randomized polynomial time algorithm that asks $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries, and finds all edges with probability $1 - O(1/m^{0.02})$.*

Theorem 1.1 follows from the theorem and (6).

For the first algorithm, let $\delta = 0.05$ and assume that the maximum degree of G is less than $m^{2\delta}$. To present the algorithm, construct a random partition $X_1, \dots, X_{m^{1/2+2\delta}}$ of X so that each vertex $x \in X$ is equally likely in X_j , $j = 1, \dots, m^{1/2+2\delta}$, independently of all other vertices. Similarly, construct a random partition $Y_1, \dots, Y_{m^{1/2+2\delta}}$ of Y .

Lemma 4.3. *Under the same hypotheses as in Theorem 4.2, if the maximum degree of G is less than $m^{2\delta}$, then, with probability $1 - (1 + o(1))m^{-\delta}$, the followings hold.*

(a) For each $i = 1, \dots, m^{1/2+2\delta}$, $|N(X_i)| \leq 2m^{1/2-2\delta}$, where $N(X_i) := \{y \in Y : y \sim x \text{ for some } x \in X_i\}$.

- (b) For each $i = 1, \dots, m^{1/2+2\delta}$ and $y \in Y$, $d(y; X_i) \leq 3$, where $d(y; X_i) := \{x \in X_i : x \sim y\}$.
- (c) For each $i = 1, \dots, m^{1/2+2\delta}$, the number of vertices $y \in Y$ with $d(y; X_i) \geq 2$ is at most $m^{5\delta}$.
- (d) The statements (a)-(c) hold after the roles of X and Y are switched.
- (e) Except for $3m^{1-3\delta}$ edges, every edge is a unique edge between X_i and Y_j for some pair i, j .

Proof. Let $p = m^{-1/2-2\delta}$. Then, $\Pr[x \in X_i] = p$ for all x and i . It is enough to show that (a)-(c) hold with probability $1 - o(m^{-\delta})$ and (e) holds with probability $1 - m^{-\delta}$.

For (a), as

$$E[|N(X_i)|] = \sum_{y \in Y} \left(1 - \Pr[X_i \cap N(y) = \emptyset]\right) = \sum_{y \in Y} \left(1 - (1-p)^{d(y)}\right) \leq \sum_{y \in Y} pd(y) \leq pm = m^{1/2-2\delta},$$

the generalized martingale inequality (Lemma 2.4) with $p = m^{-1/2-2\delta}$, $c_x = d(x)$, $\lambda = m^{1/2-2\delta}$, and $\rho = m^{-2\delta}/2$, gives that

$$\Pr\left[|N_H(X_i)| \geq 2m^{1/2-2\delta}\right] \leq 2 \exp\left(-\frac{m^{1/2-4\delta}}{2} + \frac{m^{-1/2-6\delta}}{8} \sum_{x \in X} (d(x))^2 e^{m^{-2\delta}d(x)/2}\right).$$

Since $e^{m^{-2\delta}d(x)/2} \leq e^{1/2} \leq 2$ and $\sum_{x \in X} (d(x))^2 \leq m^{2\delta} \sum_{x \in X} d(x) = m^{1+2\delta}$, we have

$$\Pr\left[|N(X_i)| \geq 2m^{1/2-2\delta}\right] \leq 2 \exp\left(-\frac{m^{1/2-4\delta}}{4}\right),$$

and

$$\Pr\left[\exists i \text{ s.t. } |N(X_i)| \geq 2m^{1/2-2\delta}\right] \leq 2m^{1/2+2\delta} \exp\left(-\frac{m^{1/2-4\delta}}{4}\right) = o(m^{-\delta}).$$

For (b),

$$\Pr[d(y; X_i) \geq 4] \leq \binom{d(y)}{4} p^4 \leq \frac{(pd(y))^4}{24}.$$

Thus, the probability that there is a pair y, j such that $d(y, X_j) \geq 4$ is at most

$$\sum_{j=1}^{m^{1/2+2\delta}} \sum_{y \in Y} \frac{(pd(y))^4}{24} \leq \frac{p^4 m^{1/2+2\delta} m^{6\delta}}{24} \sum_{y \in Y} d(y) \leq \frac{m^{-2-8\delta} m^{1/2+2\delta} m^{1+6\delta}}{24} = \frac{1}{24m^{1/2}} = o(m^{-\delta}).$$

For (c), suppose the number Z_i of vertices $y \in Y$ with $d(y, X_i) \geq 2$ is more than $m^{5\delta}$. Then there are distinct vertices y_1, \dots, y_{m^δ} in Y with $d(y_j, X_i) \geq 2$, $j = 1, \dots, m^\delta$, such that $N(y_j) \cap N(y_k) = \emptyset$ for all distinct pairs $j, k = 1, \dots, m^\delta$. This is possible since each fixed $y \in Y$ satisfies $N(y) \cap N(y') \neq \emptyset$ for at most $m^{4\delta} - 1$ vertices $y' \in Y$. As $r! \geq (\frac{r}{e})^r$ and $(d(y_j))^2 \leq m^{2\delta}d(y_j)$,

$$\Pr[Z_i > m^{5\delta}] \leq \frac{1}{m^{\delta!}} \sum_{y_1, \dots, y_{m^\delta}} \prod_{j=1}^{m^\delta} p^2 \binom{d(y_j)}{2} \leq \left(\frac{ep^2 m^{2\delta}}{2m^\delta}\right)^{m^\delta} \sum_{y_1, \dots, y_{m^\delta}} \prod_{j=1}^{m^\delta} d(y_j) \leq \left(\frac{ep^2 m^\delta m}{2}\right)^{m^\delta}$$

and

$$\Pr[\exists i \text{ s.t. } Z_i > m^{5\delta}] \leq m^{1/2+2\delta} \left(\frac{e}{2m^{3\delta}}\right)^{m^\delta} = o(m^{-\delta}).$$

For (e), the probability that an edge $e = xy$ is not a unique edge between any pair of X_i and Y_j is

$$\sum_{i,j=1}^{m^{1/2+2\delta}} \Pr[(x, y) \in X_i \times Y_j] \Pr\left[\exists \text{ edge between } X_i \text{ and } Y_j \text{ other than } e \mid (x, y) \in X_i \times Y_j\right].$$

Since the conditional probability is at most

$$(d(x) - 1)p + (d(y) - 1)p + (m - d(x) - d(y) + 1)p^2 \leq 2m^{2\delta}p + mp^2 \leq 3m^{-4\delta}$$

and $\sum_{i,j=1}^{m^{1/2+2\delta}} \Pr[(x, y) \in X_i \times Y_j] = 1$, the number W of edges that are not a unique edge between any pair of X_i and Y_j is at most $3m^{1-4\delta}$ in expectation. Markov inequality implies that

$$\Pr[W \geq 3m^{1-3\delta}] \leq m^{-\delta}.$$

□

The next algorithm finds all edges of G when the maximum degree of G is less than $m^{2\delta}$.

Algorithm A (i) For each i , $i = 1, \dots, m^{1/2+2\delta}$, regarding each $y \in Y$ as a coin with weight $w_i(y) := w_G(X_i, y) = \sum_{x \in X_i} w_G(x, y)$, apply the coin weighing algorithm in Corollary 1.3 to find all counterfeit coins with parameters $(m, n, \alpha, \beta, \varepsilon, \mu)$ replaced by $(2m^{1/2-2\delta}, n, \alpha, 3\beta, m^{-1/2+7\delta}, \frac{4}{1-4\delta})$. Let $N_0(X_i)$ be the set of all counterfeit coins found, $i = 1, \dots, m^{1/2+2\delta}$. Do the same for Y_j and let $N_0(Y_j)$ be the set of all counterfeit coins found, $j = 1, \dots, m^{1/2+2\delta}$.

(ii) For all pairs $i, j = 1, \dots, m^{1/2+2\delta}$ with $|N_0(X_i) \cap Y_j| = |X_i \cap N_0(Y_j)| = 1$, take $y \in N_0(X_i) \cap Y_j$ and $x \in X_i \cap N_0(Y_j)$ and weigh the possible edge xy to obtain $w_G(x, y)$. For each pair xy with $w_G(x, y) \neq 0$, declare that xy is an edge of G .

(iii) Find remaining edges one by one by applying the randomized binary search using no more than $(6\lceil \log n \rceil + 15)m^{1-3\delta}$ queries.

For the collectedness and the query complexity of the algorithm, we prove the following lemma.

Lemma 4.4. *Under the same hypotheses as in Theorem 4.2, if the maximum degree of G is less than $m^{2\delta}$, then, with probability $1 - (1 + o(1))m^{-\delta}$, Algorithm A asks $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries to find all edges of G .*

Proof. Suppose (a)-(e) of Lemma 4.5 hold. First, we show that the parameters $(2m^{1/2-2\delta}, n, \alpha, 3\beta, m^{-1/2+7\delta})$ satisfy all the requirements in Corollary 1.3. If y is counterfeit, then $w_i(y) = w_G(X_i, y) \neq 0$. This gives $y \in N(X_i)$ and hence the number of counterfeit coins is at most $|N(X_i)| \leq 2m^{1/2-2\delta}$ by (a) of Lemma 4.3. The number of all coins is $|Y| \leq n$. If $y \sim x$ for only one $x \in X_i$, then $|w_i(y)| = |w_G(x, y)| \geq \alpha$. Thus, $0 < |w_i(y)| < \alpha$ implies $d(y; X_i) \geq 2$. The number of such $y \in Y$ is at most $m^{5\delta} = m^{-1/2+7\delta} \cdot 2m^{1/2-2\delta}$ by (c) of Lemma 4.3. Since $d(y; X_i) \leq 3$ by (b) of Lemma 4.3, $|w_i(y)| \leq \sum_{x \in X_i} |w(x, y)| \leq 3\beta$. Therefore, the algorithm finds the set $N_0(X_i)$ of all counterfeit coins, with probability $1 - O(m^{-2})$ for each X_i . Similarly, the algorithm finds the set $N_0(Y_j)$ of all counterfeit coins, with probability $1 - O(m^{-2})$ for each Y_j . Since there are $2m^{1/2+2\delta}$ sets X_i and Y_j , $N_0(X_i) = \{y \in Y : w_i(y) \neq 0\}$ and $N_0(Y_j) = \{x \in X : w_j(x) \neq 0\}$, with probability $1 - O(1/m)$.

If $e = xy$ is a unique edge between X_i and Y_j , then $|w_i(y)|, |w_j(x)| \geq \alpha$, especially, $y \in N_0(X_i)$ and $x \in N_0(Y_j)$. Moreover, as there is no other edge between X_i and Y_j , $N_0(X_i) \cap Y_j = \{y\}$ and $X_i \cap N_0(Y_j) = \{x\}$. Thus, the algorithm finds the edge $e = xy$ in (ii). By (e) of Lemma 4.3, at most $3m^{1-3\delta}$ edges remain unfound in (ii). All the remaining edges can be found in (iii) with probability $1 - e^{-\Omega(m^{1-3\delta})}$ by Lemma 3.1.

For the query complexity, in (i), $O(\frac{m^{1/2-2\delta} \log(\beta/\alpha) \log n}{\log m})$ queries are enough for each X_i or Y_j . As there are $2m^{1/2+2\delta}$ such sets, $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries are enough in (i). In (ii), if $|N_0(X_i) \cap Y_j| = |X_i \cap N_0(Y_j)| = 1$, then there is at least one edge between X_i and Y_j . As there are at most m such pairs X_i, Y_j , m queries are enough in (ii). Since $o(\frac{m \log n}{\log m})$ queries are asked in (iii), the query complexity of the algorithm is $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$.

□

For general graphs, select each vertex of Y with probability $m^{-\delta}$, $\delta = 0.05$, independently of all other vertices. Let G_1 be the induced graph on X and the selected vertices of Y .

Lemma 4.5. *If G has at most m edges, then the followings hold with probability $1 - O(m^{-\delta/2})$.*

- (a) *The number of edges in G_1 is at most $m^{1-\delta/2}$.*
- (b) *If $d_{G_1}(x) \geq m^\delta/2$, then $d_G(x) \leq 2m^\delta d_{G_1}(x) \leq 3d_G(x)$.*
- (c) *If $d_G(x) \geq m^{2\delta}$, then $d_{G_1}(x) \geq m^\delta/2$.*

Proof. As each edge in G_1 with probability $m^{-\delta}$, the expected number of edges in G_1 is at most $m^{1-\delta}$. Markov Inequality gives

$$\Pr[\text{the number of edges in } G_1 \geq m^{1-\delta/2}] \leq m^{-\delta/2}.$$

For the degree $d_{G_1}(x)$ of x in G_1 , as $E[d_{G_1}(x)] = m^{-\delta}d_G(x)$, Lemma 2.4 with $c_y = 1$ if $y \sim x$ and $c_y = 0$ otherwise, $\lambda = \frac{m^\delta}{4}$, $\rho = 1/2$ gives, for $x \in X$ with $d_G(x) < \frac{m^{2\delta}}{4}$,

$$\Pr\left[|d_{G_1}(x) - m^{-\delta}d_G(x)| \geq \frac{m^\delta}{4}\right] \leq 2 \exp\left(-\frac{m^\delta}{8} + \frac{e^{1/2}m^{-\delta}d_G(x)}{8}\right) \leq 2 \exp\left(-\frac{m^\delta}{16}\right).$$

In particular, if $d_G(x) < \frac{m^{2\delta}}{4}$, then $d_{G_1}(x) - m^{-\delta}d_G(x) < m^\delta/4$, or equivalently, $d_{G_1}(x) < m^\delta/4 + m^{-\delta}d_G(x) < m^\delta/2$, with probability $1 - e^{-\Omega(m^\delta)}$.

For (b), it is now enough to show that $d_G(x) \leq 2m^\delta d_{G_1}(x) \leq 3d_G(x)$ when $d_G(x) \geq \frac{m^{2\delta}}{4}$, say, with probability $1 - e^{-\Omega(m^\delta)}$. Lemma 2.4 with $c_y = 1$ if $y \sim x$ and $c_y = 0$ otherwise, $\lambda = \frac{m^{-\delta}d_G(x)}{2}$, $\rho = 1/3$ also gives

$$\Pr\left[|d_{G_1}(x) - m^{-\delta}d_G(x)| \geq \frac{m^{-\delta}d_G(x)}{2}\right] \leq 2 \exp\left(-\frac{m^{-\delta}d_G(x)}{6} + \frac{e^{1/3}m^{-\delta}d_G(x)}{18}\right) \leq 2 \exp\left(-\frac{m^{-\delta}d_G(x)}{12}\right),$$

for $e^{1/3} \leq 3/2$. If $d_G(x) \geq m^{2\delta}/4$, we have $|2m^\delta d_{G_1}(x) - 2d_G(x)| \leq d_G(x)$, or equivalently, $d_G(x) \leq 2m^\delta d_{G_1}(x) \leq 3d_G(x)$, with probability $1 - e^{-\Omega(m^\delta)}$. Moreover, if $d_G(x) \geq m^{2\delta}$, then $2m^\delta d_{G_1}(x) \geq d_G(x) \geq m^{2\delta}$. That is, $d_{G_1}(x) \geq m^\delta/2$, which shows (c). □

Algorithm B (i) Apply the randomized binary search to find edges of G_1 one by one, using $(2\lceil \log n \rceil + 5)m^{1-\delta/2}$ queries. Let G_2 be the graph on $X \cup Y$ consisting of all edges found.

(ii) For each vertex $x \in X$ with $d_{G_2}(x) \geq m^\delta/2$, regard each $y \in Y$ as a coin with weight $w_x(y) := w_G(x, y)$ and apply the coin weighing algorithm in Corollary 1.3 with parameters $(m, n, \alpha, \beta, \varepsilon, \mu)$ replaced by $(2m^\delta d_{G_2}(x), n, \alpha, \beta, 0, 1/\delta)$. The vertices $x \in X$ with $d_{G_2}(x) \geq m^\delta/2$ are called vertices of large degree.

(iii) Output vertices of large degree and all edges found.

Algorithm B has the following property.

Lemma 4.6. *Under the same hypotheses as in Theorem 4.2, with probability $1 - O(m^{-\delta/2})$, Algorithm B uses $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries to find all vertices $x \in X$ with $d_G(x) \geq m^{2\delta}$ and all edges containing them.*

Proof. Suppose (a) and (b) of Lemma 4.5 hold. Then Lemma 4.1 yields $G_2 = G_1$ with probability $1 - e^{-\Omega(m^{1-\delta/2})}$. We assume that $G_1 = G_2$ in the rest of the proof.

In (ii), note that the number of counterfeit coins for x is $d_G(x)$, which is at most $2m^\delta d_{G_2}(x)$ for all $d_{G_1}(x) = d_{G_2}(x) \geq m^\delta/2$ by (b) of Lemma 4.5. Thus, the algorithm in Corollary 1.3 finds $N_G(x)$ for each $x \in X$ satisfying $d_{G_2}(x) \geq m^\delta/2$, with probability $1 - O(1/(2m^\delta d_{G_2}(x))^{1/\delta}) = 1 - O(1/m^2)$. As $d_G(x) \geq d_{G_2}(x)$, there are at most $2m^{1-\delta}$ vertices $x \in X$ with $d_{G_2}(x) \geq m^\delta/2$ and the algorithm finds $N_G(x)$ for all such vertices $x \in X$, with probability $1 - O(1/m)$. In particular, if $d_G(x) \geq m^{2\delta}$, then $d_{G_2}(x) \geq m^\delta/2$ by (c) of Lemma 4.5 and hence $N_G(x)$ are found.

For the query complexity, $(2\lceil \log n \rceil + 5)m^{1-\delta/2}$ queries are asked in (i). In (ii), $O(\frac{m^\delta d_{G_2}(x) \log(\beta/\alpha) \log n}{\log m})$ queries are asked for each $x \in X$ with $d_{G_2}(x) \geq m^\delta/2$. On the other hand, $d_{G_2}(x) \geq m^\delta/2$ implies $2m^\delta d_{G_2}(x) \leq 3d_G(x)$ by (b) of Lemma 4.5. Thus,

$$\sum_{x: d_{G_2}(x) \geq m^\delta/2} m^\delta d_{G_2}(x) \leq \frac{3}{2} \sum_{x \in X} d_G(x) = \frac{3m}{2}$$

gives that $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries are asked in (ii). □

To find all vertices v in G with $d_G(v) \geq m^{2\delta}$, one may apply Algorithm B twice, one as it is and the other after exchanging roles of X and Y . Then, after removing all vertices found (and all edges containing any of them), we apply Algorithm A. Lemmas 4.4 and 4.6 imply that

Corollary 4.7. *Under the same hypotheses as in Theorem 4.2, there is a polynomial time randomized algorithm asking $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries to find all edges of G , with probability $1 - O(1/m^{0.02})$.*

If the algorithm in the corollary is forced to stop when it asks $\frac{\eta m \log(\beta/\alpha) \log n}{\log m}$ queries, for the constant η in the $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ term, the desired algorithm in Theorem 4.2 may be obtained.

5 Concluding Remarks

In this paper, we presented a polynomial time randomized algorithm that uses $O(\frac{m \log(\beta/\alpha) \log n}{\log m})$ queries, when there are at most m counterfeit coins and the weights $w(c)$ of all counterfeit coins satisfy $\alpha \leq |w(c)| \leq \beta$. This plays a key role to find a hidden weighted graph G satisfying similar conditions. Though there is a non-adaptive algorithm to find all counterfeit coins using $O(\frac{m \log n}{\log m})$ queries [11], it is not a polynomial time algorithm. An obvious question is if there is a polynomial time algorithm to find all counterfeit coins using $O(\frac{m \log n}{\log m})$ queries when there is no restriction on the weights.

The algorithm we presented was a randomized algorithm that uses the optimal number of queries up to a constant factor. On the other hand, the best deterministic algorithm uses $\Theta(\frac{m \log n}{\log m} + m \log \log m)$ (see [9]), it would be good to implement a deterministic polynomial time algorithm that uses $O(\frac{m \log n}{\log m})$ queries even when the weights of counterfeit coins are positive real numbers.

References

- [1] N. Alon and V. Asodi. Learning a hidden subgraph. *SIAM Journal on Discrete Mathematics*, 18(4):697–712, 2005.
- [2] N. Alon, R. Beigel, S. Kasif, S. Rudich, and B. Sudakov. Learning a hidden matching. *SIAM Journal on Computing*, 33(2):487–501, 2004.
- [3] D. Angluin and J. Chen. Learning a hidden graph using $O(\log n)$ queries per edge. In *Proceedings of the 17th Annual Conference on Learning Theory (COLT 2004)*, pages 210–223, Banff, Canada, 2004.
- [4] D. Angluin and J. Chen. Learning a hidden hypergraph. *Journal of Machine Learning Research*, 7:2215–2236, 2006.
- [5] R. Beigel, M. S. Apaydin, L. Fortnow, and S. Kasif. An optimal procedure for gap closing in whole genome shotgun sequencing. In *Proceedings of the Fifth Annual International Conference on Computational Molecular Biology (RECOMB 2001)*, pages 22–30, 2001.

- [6] M. Bouvel, V. Grebinski, and G. Kucherov. Combinatorial search on graphs motivated by bioinformatics applications: A brief survey. In *the 31st International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2005)*, pages 16–27, 2005.
- [7] N. H. Bshouty. Optimal algorithms for the coin weighing problem with a spring scale. In *Proceedings of the 22nd Annual Conference on Learning Theory (COLT 2009)*, Montreal, Canada, 2009.
- [8] N. H. Bshouty and H. Mazzawi. Optimal query complexity for reconstructing hypergraphs. In *Proceedings of the 27th International Symposium on Theoretical Aspects of Computer Science (STACS 2010)*, pages 143–154, Nancy, France, 2010.
- [9] N. H. Bshouty and H. Mazzawi. Toward a deterministic polynomial time algorithm with optimal additive query complexity. In *Proceedings of the 35th International Symposium on Mathematical Foundations of Computer Science (MFCS 2010)*, pages 221–232, Brno, Czech Republic, 2010.
- [10] N. H. Bshouty and H. Mazzawi. On parity check $(0, 1)$ -matrix over \mathbb{Z}_p . In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA 2011)*, pages 1383–1394, San Francisco, USA, 2011.
- [11] N. H. Bshouty and H. Mazzawi. Reconstructing weighted graphs with minimal query complexity. *Theoretical Computer Science*, 412(19):1782–1790, 2011.
- [12] D. G. Cantor. Determining a set from the cardinalities of its intersections with other sets. *Canadian Journal of Mathematics*, 16:94–97, 1964.
- [13] D. G. Cantor and W. H. Mills. Determination of a subset from certain combinatorial properties. *Canadian Journal of Mathematics*, 18:42–48, 1966.
- [14] J. Capetanakis. Generalized TDMA: The multi-accessing tree protocol. *IEEE Transactions on Communications*, 27(10):1476–1484, 1979.
- [15] J. Capetanakis. Tree algorithms for packet broadcast channels. *IEEE Transactions on Information Theory*, 25(5):505–515, 1979.
- [16] S. S. Choi, K. Jung, and J. H. Kim. Almost tight upper bound for finding Fourier coefficients of k -bounded pseudo-Boolean functions. *Journal of Computer and System Sciences*, 77(6):1039–1053, 2011.
- [17] S. S. Choi and J. H. Kim. Randomized polynomial time algorithms for finding weighted graphs with optimal additive query complexity. submitted.
- [18] S. S. Choi and J. H. Kim. Optimal query complexity bounds for finding graphs. *Artificial Intelligence*, 174(9–10):551–569, 2010.
- [19] S. S. Choi and J. H. Kim. Sample complexity for linkage learning. Submitted, 2011.
- [20] D. Du and F. K. Hwang. Combinatorial group testing and its application. In *V. 3 of Series on applied mathematics*, chapter 10. World Science, 1993.
- [21] P. Erdős and A. Rényi. On two problems of information theory. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 8:241–254, 1963.
- [22] N. J. Fine. Solution of problem E 1399. *American Mathematical Monthly*, 67(7):697–698, 1960.
- [23] V. Grebinski. On the power of additive combinatorial search model. In *Proceedings of the 4th Annual International Conference on Computing and Combinatorics (COCOON 1998)*, pages 194–203, Taipei, Taiwan, 1998.

- [24] V. Grebinski and G. Kucherov. Optimal query bounds for reconstructing a Hamiltonian cycle in complete graphs. In *the Fifth Israel Symposium on the Theory of Computing Systems (ISTCS 1997)*, pages 166–173, 1997.
- [25] V. Grebinski and G. Kucherov. Reconstructing a Hamiltonian cycle by querying the graph: Application to DNA physical mapping. *Discrete Applied Mathematics*, 88:147–165, 1998.
- [26] V. Grebinski and G. Kucherov. Optimal reconstruction of graphs under the additive model. *Algorithmica*, 28:104–124, 2000.
- [27] J. J. Hein. An optimal algorithm to reconstruct trees from additive distance data. *Bulletin of Mathematical Biology*, 51(5):597–603, 1989.
- [28] J. H. Kim. The Ramsey number $R(3, t)$ has order of magnitude $t^2/\log t$. *Random Structures and Algorithms*, 7(3):173–208, 1995.
- [29] V. King, L. Zhang, and Y. Zhou. On the complexity of distance-based evolutionary tree reconstruction. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2003)*, pages 444–453, 2003.
- [30] M. Li and P. M. B. Vitányi. Kolmogorov complexity arguments in combinatorics. *J. Comb. Theory Series A*, 66(2):226–236, 1994.
- [31] B. Lindström. On a combinatorial detection problem I. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 9:195–207, 1964.
- [32] B. Lindström. On a combinatorial problem in number theory. *Canadian Mathematical Bulletin*, 8(4):477–490, 1965.
- [33] B. Lindström. On Möbius functions and a problem in combinatorial number theory. *Canadian Mathematical Bulletin*, 14(4):513–516, 1971.
- [34] B. Lindström. Determining subsets by unramified experiments. In J. N. Srivastava, editor, *A Survey of Statistical Designs and Linear Models*, pages 407–418. North Holland, 1975.
- [35] J. L. Massey. Collision-resolution algorithms and random-access communications. In G. Longo, editor, *Multi-user communications systems, CISM Courses and Lecture Notes No. 265*, pages 73–137. Springer, Wien and New York, 1981.
- [36] H. Mazzawi. Optimally reconstructing weighted graphs using queries. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA 2010)*, pages 608–615, Austin, USA, 2010.
- [37] C. McDiarmid. On the method of bounded differences. In J. Siemons, editor, *Surveys in Combinatorics*, London Mathematical Society Lecture Note Series 141, pages 148–188. Cambridge University Press, 1989.
- [38] L. Moser. The second moment method in combinatorial analysis. In *Combinatorial Structures and Their Applications. Proceedings of the Calgary International Conference on Combinatorial Structures and Their Applications held at the University of Calgary. June 1969*, pages 283–384. Gordon and Breach, New York, 1970.
- [39] L. Reyzin and N. Srivastava. Learning and verifying graphs using queries with a focus on edge counting. In *Proceedings of the 18th International Conference on Algorithmic Learning Theory (ALT 2007)*, pages 285–297, Sendai, Japan, 2007.
- [40] L. Reyzin and N. Srivastava. On the longest path algorithm for reconstructing trees from distance matrices. *Information Processing Letters*, 101(3):98–100, 2007.

- [41] H. S. Shapiro. Problem E 1399. *American Mathematical Monthly*, 67(1):82, 1960.
- [42] S. Söderberg and H. S. Shapiro. A combinatory detection problem. *American Mathematical Monthly*, 70:1066–1070, 1963.
- [43] H. Tettelin, D. Radune, S. Kasif, H. Khouri, and S. L. Salzberg. Optimized multiplex PCR: Efficiently closing a whole-genome shotgun sequencing project. *Genomics*, 62:500–507, 1999.
- [44] B. Tsybakov and V. Mikhailov. Free synchronous packet access in a broadcast channel with feedback. *Problemy Peredachi Informassi*, 14(4):259–280, 1978.
- [45] R. Uehara, K. Tsuchida, and I. Wegener. Identification of partial disjunction, parity, and threshold functions. *Theoretical Computer Science*, 210(1–2):131–147, 2000.

Appendix: Proofs of Lemmas 3.1, 3.2 and 4.1

In this appendix, we prove Lemmas 3.1, 3.2 and 4.1.

Lemma 3.1 *The randomized binary search finds all counterfeit coins with probability $1 - e^{-\Omega(m)}$.*

Proof. If there is a counterfeit coin c , conditioned on $A'' = A' \setminus \{c\}$, A' can be one of A'' and $A'' \cup \{c\}$, each with probability $1/2$. Since $w(A'' \cup \{c\}) = w(A'') + w(c) \neq w(A'')$, the probability of $w(A') \neq 0$ is at least $1/2$.

Let Z_i be the number of random trials when the i^{th} counterfeit coin is found. Then, for $a = \ln(4/3)$,

$$E[e^{aZ_i} | Z_1, \dots, Z_{i-1}] = \sum_{k=1}^{\infty} e^{ak} (1 - p_i)^{k-1} p_i = \frac{p_i}{1 - p_i} \frac{(1 - p_i)e^a}{1 - (1 - p_i)e^a} = \frac{p_i e^a}{1 - (1 - p_i)e^a} \leq \frac{e^a}{2 - e^a},$$

for $p_i := \Pr[w(A') \neq 0 | Z_1, \dots, Z_{i-1}] \geq 1/2$. As

$$E[e^{a \sum_{i=1}^{\ell} Z_i}] = E\left[E\left[e^{a \sum_{i=1}^{\ell} Z_i} \middle| Z_1, \dots, Z_{\ell-1}\right]\right] = E\left[e^{a \sum_{i=1}^{\ell-1} Z_i} E\left[e^{aZ_{\ell}} \middle| Z_1, \dots, Z_{\ell-1}\right]\right] \leq \left(\frac{e^a}{2 - e^a}\right) E\left[e^{a \sum_{i=1}^{\ell-1} Z_i}\right],$$

for all $\ell = 1, \dots, m$, we have

$$\Pr\left[\sum_{i=1}^{m^*} Z_i \geq 3m\right] \leq \Pr[e^{a \sum_{i=1}^{m^*} Z_i} \geq e^{3am}] \leq E[e^{a(\sum_{i=1}^{m^*} Z_i - 3m)}] \leq \left(\frac{e^{-2a}}{2 - e^a}\right)^m \leq \left(\frac{27}{32}\right)^m = e^{-\Omega(m)},$$

where $m^* \leq m$ is the number of counterfeit coins,

□

Lemma 3.2 *Suppose a set A of n or less coins are given, and the number of counterfeit coins in A is at most $q \geq 2$. If the weights $w(c)$ of all but at most $q/2$ counterfeit coins c satisfy $|w(c)| \geq \alpha$. Then, with probability $1 - O(\frac{1}{q})$, we have the followings.*

- (a) *There are at most $\frac{5q}{6}$ counterfeit coins c that satisfy $|w(c)| < \alpha$ (not exclusively) or belong to a set $A_{0,j}$ containing more than one counterfeit coin, $j = 1, \dots, 2^{\ell_q}$.*
- (b) *For each $i = 1, \dots, \lceil 2 \log q \rceil - 1$, $A_{i,j}$ contains at most $\frac{i+2 \log q}{i}$ counterfeit coins.*
- (c) *For each $i = 1, \dots, \lceil 2 \log q \rceil - 1$, there are at most $2^{-(i+1)}q + q^{3/4}$ sets $A_{i,j}$ that contain more than one counterfeit coin.*
- (d) *For $i \geq \lceil 2 \log q \rceil - 1$, each $A_{i,j}$ contains one or less counterfeit coin.*
- (e) *Each $A_{\lceil 3 \log n \rceil, j}$ contains at most one coin.*

Proof. (a) For any counterfeit coin c , the probability that c belongs to a set $A_{0,j}$ containing another counterfeit coin is at most $1 - (1 - 2^{-\ell_q})^{q-1} \leq 1 - 1/e$. Thus, the number of such counterfeit coins c with $|w(c)| \geq \alpha$ is at most $(1 - 1/e)(q - q_1)$ in expectation, where q_1 is the number of counterfeit coins c with $|w(c)| < \alpha$. As $q - q_1 \geq q/2$ and the number depends only on where counterfeit coins are in, we may apply the Azuma-Hoeffding martingale inequality (Lemma 2.3) with $c_\ell = 2$ and $\sum_\ell c_\ell^2 \leq 4q$ to deduce that the number of such counterfeit coins c with $|w(c)| \geq \alpha$ is at most $2(q - q_1)/3$, with probability $1 - e^{-\Omega(q)}$. Thus, with probability $1 - e^{-\Omega(q)}$, there are at most

$$2(q - q_1)/3 + q_1 = 2q/3 + q_1/3 \leq 2q/3 + q/6 = 5q/6$$

counterfeit coins c that satisfy $|w(c)| < \alpha$ or belong to a set $A_{0,j}$ containing another counterfeit coin.

(b) For each set $A_{i,j}$, the probability that $A_{i,j}$ contains $k_i := \lceil \frac{i+2\log q}{i} \rceil$ or more counterfeit coins are bounded from above by

$$\binom{q}{k_i} 2^{-k_i(\ell_q+i)} \leq 2^{-k_i(\ell_q+i)} q^{k_i}.$$

Thus, for each $i = 1, \dots, \lceil 2\log q \rceil - 1$, the probability that $A_{i,j}$ contains k_i or more counterfeit coins is at most

$$2^{\ell_q+i} 2^{-k_i(\ell_q+i)} q^{k_i} = 2^{-(k_i-1)(\ell_q+i)} q^{k_i} \leq 2^{-(k_i-1)i} q \leq \frac{1}{q}.$$

(c) The probability that $A_{i,j}$ contains two or more counterfeit coins is at most

$$\binom{q}{2} 2^{-2(\ell_q+i)} \leq 2^{-2(\ell_q+i)-1} q^2,$$

and, for each i , the expected number of $A_{i,j}$ containing two or more counterfeit coins is at most

$$2^{\ell_q+i} 2^{-2(\ell_q+i)-1} q^2 \leq 2^{-(\ell_q+i)-1} q^2 = 2^{-(i+1)} q.$$

Counting coordinates corresponding to counterfeit coins only, we apply the Azuma-Hoeffding martingale inequality (Lemma 2.3) with $c_\ell = 1$ to conclude that, for each i , the number of $A_{i,j}$ containing two or more counterfeit coins is at most $2^{-(i+1)} q + q^{3/4}$, with probability $1 - e^{-\Omega(q^{1/2})}$.

(d) For $i = \lceil 2\log q \rceil - 1$, by the same estimation as in (c), the probability that $A_{i,j}$ contains two or more counterfeit coins for some j is at most $2^{-(i+1)} q \leq 1/q$. If each $A_{\lceil 2\log q \rceil - 1, j}$ contains at most one counterfeit coin, then so does each $A_{i,j}$ with $i \geq \lceil 2\log q \rceil$, for every set $A_{i,j}$ with $i \geq \lceil 2\log q \rceil$ is a subset of some $A_{\lceil 2\log q \rceil - 1, \ell}$.

(e) The statement follows since each set $A_{i,j}$ with $i \geq \lceil 2\log q \rceil - 1$ is deterministically divide into two sets with size difference at most 1.

□

Lemma 4.1 *The randomized binary search finds all edges of G with probability $1 - e^{-\Omega(m)}$.*

Proof. If G has at least one edge, say $e = xy$, conditioned $X'' := X' \setminus \{x\}$ and $Y'' = Y' \setminus \{y\}$, (X', Y') can be one of (X'', Y'') , $(X'' \cup \{x\}, Y'')$, $(X'', Y'' \cup \{y\})$, and $(X'' \cup \{x\}, Y'' \cup \{y\})$, each with probability $1/4$. If all three weights $w(X'', Y'')$, $w(X'' \cup \{x\}, Y'')$, $w(X'', Y'' \cup \{y\})$ are 0, then $w(X'' \cup \{x\}, Y'' \cup \{y\}) = w(e) \neq 0$. In other words, at least one of the four weights is non-zero. This yields that the probability of $w(X', Y') \neq 0$ is at least $1/4$.

Let Z_i be the number of random trials when the i^{th} edge is found. Then, for $a = \ln(13/12)$,

$$E[e^{aZ_i} | Z_1, \dots, Z_{i-1}] = \sum_{k=1}^{\infty} e^{ak} (1 - p_i)^{k-1} p_i = \frac{p_i}{1 - p_i} \frac{(1 - p_i)e^a}{1 - (1 - p_i)e^a} = \frac{p_i e^a}{1 - (1 - p_i)e^a} \leq \frac{e^a}{4 - 3e^a},$$

for $p_i := \Pr[w(X', Y') \neq 0 | Z_1, \dots, Z_{i-1}] \geq 1/4$. Thus,

$$\Pr\left[\sum_{i=1}^{m^*} Z_i \geq 5m\right] \leq \Pr[e^{a \sum_{i=1}^{m^*} Z_i} \geq e^{5am}] \leq E[e^{a(\sum_{i=1}^{m^*} Z_i - 5m)}] \leq \left(\frac{e^{-4a}}{4 - 3e^a}\right)^m \leq (0.97)^m = e^{-\Omega(m)},$$

where $m^* \leq m$ is the number of edges in G , as

$$E\left[e^{a \sum_{i=1}^{\ell} Z_i}\right] = E\left[E\left[e^{a \sum_{i=1}^{\ell} Z_i} \middle| Z_1, \dots, Z_{\ell-1}\right]\right] = E\left[e^{a \sum_{i=1}^{\ell-1} Z_i} E\left[e^{a Z_{\ell}} \middle| Z_1, \dots, Z_{\ell-1}\right]\right] \leq \left(\frac{e^a}{4 - 3e^a}\right) E\left[e^{a \sum_{i=1}^{\ell-1} Z_i}\right],$$

for all $\ell = 1, \dots, m^*$.

□